



Does education reduce wage inequality? Quantile regression evidence from 16 countries

Pedro S. Martins^{a,*}, Pedro T. Pereira^b

^a*Department of Economics, University of Warwick, Coventry CV47AL, UK*

^b*Universidade da Madeira, CEPR and IZA, UK*

Received 31 July 2001; received in revised form 29 January 2003; accepted 30 May 2003

Abstract

Quantile regression estimates of returns to education are used to address the relation between schooling and wage inequality. Empirical evidence for male workers from 16 countries for the mid-1990s suggests a robust stylised fact: Returns to schooling are higher for the more skilled individuals, conditional on their observable characteristics. This suggests that schooling has a positive impact upon within-levels wage inequality. Factors such as over-education, ability–schooling interactions and school quality or different fields of study may be driving this result.
© 2003 Elsevier B.V. All rights reserved.

JEL classification: C29; J31; I21

Keywords: Returns to education; Wage inequality; Quantile regression

1. Introduction

Returns to education have been thoroughly analysed in the labour economics literature. This is understandable as the pay-off to schooling is important information for both public and private decisions on how much to invest in the sector (see Card, 1999 for a survey).

A related issue concerns wage inequality. After witnessing major increases in the spread of wages since the early 1980s,¹ some Western decision-makers have portrayed schooling

* Corresponding author. Tel.: +44-2476-523-935; fax: +44-2476-523-032.

E-mail address: p.martins@warwick.ac.uk (P.S. Martins).

¹ See Katz and Murphy (1992) and Juhn et al. (1993) for a description of the US case and some tentative explanations.

as the best tool to erode the supposedly globalisation-related forces that increase wage inequality. As Ashenfelter and Rouse (2000, p. 111) put it, “The school is a promising place to increase the skills and incomes of individuals. As a result, educational policies have the potential to decrease existing, and growing, inequalities in income”.

This line of thought carries with it the presumption that new highly educated cohorts will benefit from such levels’ traditionally high returns. However, this approach disregards whether such levels are characterised by reasonably concentrated or disperse returns. If the latter situation turns out to be the most representative, then one should acknowledge the potential problems concerning within-levels inequality of educational policies designed to erode wage dispersion. Moreover, the scarce evidence available suggests that “differences in the extent of earnings inequality among high-income countries are heavily influenced by the rewards for educational attainment” (Sullivan and Smeeding, 1997).

Our aim with this paper is to fill in this gap by drawing on quantile regression estimates of returns to education. This approach allows us to assess the differences in the schooling-related pay increment across the wage distribution. We therefore compare the returns to education for the “skilled” and the “unskilled” workers (conditional on their schooling and experience) in order to shed light on the contribution of schooling upon within-levels wage inequality.

Furthermore, we provide evidence on this matter for a large number of countries. This was achieved under the framework of a research project, ‘Public Funding and Private Returns to Education’ (PuRE), where each country team analysed their country data sets. Special care was taken to assure that these data sources were as similar and thus comparable as possible.

The paper is structured as follows. Section 2 outlines the quantile regression methodology. The following section describes the data sets used and provides comparable descriptive statistics for the 16 countries analysed. Section 4 describes the results obtained and the following section discusses them. Finally, Section 6 concludes.

2. Quantile regression

An ordinary least squares (OLS) regression is based on the *mean* of the conditional distribution of the regression’s dependent variable. This approach is used because one implicitly assumes that possible differences in terms of the impact of the exogenous variables along the conditional distribution are unimportant.

However, this may prove inadequate in some research agendas. If exogenous variables influence parameters of the conditional distribution of the dependent variable other than the mean, then an analysis that disregards this possibility will be severely weakened (see Koenker and Bassett, 1978). Unlike OLS, quantile regression models allow for a full characterisation of the conditional distribution of the dependent variable.²

² See Abadie et al. (2002) for a recent extension of quantile regressions, considering instrumental variables.

In a wage equation setting, the quantile regression model can be written as:

$$\ln w_i = x_i \beta_\theta + u_{\theta i} \quad \text{with} \quad \text{Quant}_\theta(\ln w_i | x_i) = x_i \beta_\theta \quad (1)$$

where x_i is the vector of exogenous variables and β_θ is the vector of parameters. $\text{Quant}_\theta(\ln w | x)$ denotes the θ th conditional quantile of $\ln w$ given x . The θ th regression quantile, $0 < \theta < 1$, is defined as a solution to the problem:

$$\min_{\beta \in R^k} \left\{ \sum_{i: \ln w_i \geq x_i \beta} \theta |\ln w_i - x_i \beta_\theta| + \sum_{i: \ln w_i < x_i \beta} (1 - \theta) |\ln w_i - x_i \beta_\theta| \right\} \quad (2)$$

This is normally written as:

$$\min_{\beta \in R^k} \sum_i \rho_\theta(\ln w_i - x_i \beta_\theta), \quad (3)$$

where $\rho_\theta(\varepsilon)$ is the check function defined as $\rho_\theta(\varepsilon) = \theta \varepsilon$ if $\varepsilon \geq 0$ or $\rho_\theta(\varepsilon) = (\theta - 1)\varepsilon$ if $\varepsilon < 0$.

This problem does not have an explicit form but can be solved by linear programming methods. Standard errors are obtainable by bootstrap methods.

The least absolute deviation (LAD) estimator of β is a particular case within this framework. This is obtained by setting $\theta = 0.5$ (the median regression). The first quartile is obtained by setting $\theta = 0.25$ and so on. As one increases θ from 0 to 1, one traces the entire distribution of y , conditional on x .

Summing up, quantile regressions provide snapshots of different points of a conditional distribution. They therefore constitute a parsimonious way of describing the whole distribution and should bring much value-added if the relationship between the regressors and the independent variable evolves across its conditional distribution.

This flexibility has so far been precluded in the returns-to-education literature. In so doing, it has left unaddressed the possible impact of schooling upon inequality, through its within-levels inequality component. If the schooling-related earnings increment were the same across the wage distribution, then schooling would not impact upon within-levels wage inequality as distributions of wages conditional on different levels of schooling would differ only on their locations and not on their dispersions.

However, it may be the case that these dispersions do indeed vary across educational levels, thus resulting in an impact of schooling upon the wage distribution, through its within-levels channel. This is the possibility we test, by using quantile regression estimates, in the next two sections.

3. Data-sets description

The results for each country considered here were derived from a specific cross-section data set used by each country's team, within the above-mentioned "PuRE" research project. Table 1 describes such data sets, referring the year for which the information applies and also the number of observations used. In Appendix A, we provide a more thorough description of these data sources.

Most data sets are household surveys. The exceptions are administrative registers (the case of Denmark), labour-market surveys (France) and employer-based data sets (Nether-

Table 1
Data-sets description, descriptive statistics and inequality measures

Country	Data set	Year	No. of observations	Educ.	Exp.	Log Wage		Wage			Wage ratios (1)		
						Mean	C.V.	10th percentile	50th percentile	90th percentile	9/1	9/5	5/1
Austria	Mikrozensus	1993	7175	10.1	21.3	4.57	0.077	65.8	93.8	150	2.28	1.6	1.43
Denmark	Long. Lab. Market Reg.	1995	4416	12	19.4	4.97	0.072	96.5	138.4	230.4	2.39	1.67	1.43
Finland	Labour Force Survey	1993	1175	11.4	19.5	4.16	0.091	41.9	62.1	106.1	2.53	1.71	1.48
France	Training Qualif. + Employment Survey	1993	4606	11.4	21.9	10.92	0.036	19.8	29.8	54.1	2.73	1.81	1.5
Germany	Socio-Economic Panel	1995	1070	11.9	24.7	3.4	0.103	2.64	2.92	3.01	1.45	1.09	1.33
Greece	Household Budget Survey	1994	2096	10.1	21.9	6.93	0.092	527	1103	1907	3.62	1.73	2.09
Ireland	ESRI Household Survey	1994	1903	12.4	23.8	1.74	0.351	2.5	5.9	11.9	4.74	2.01	2.36
Italy	Survey of Household Income and Wealth	1995	3441	10.1	22.9	2.52	0.163	7.8	12.5	20.8	2.67	1.67	1.6
Netherlands	Structure of Earnings Survey	1996	49805	12.5	20	3.23	0.142	15.5	24.9	43.8	2.83	1.75	1.61
Norway	Level of Living Survey	1995	870	12.2	20.9	4.65	0.071	71.4	101.1	158	2.21	1.56	1.42
Portugal	Personnel Records	1995	28055	6.5	24.5	6.42	0.095	318	531	1456	4.58	2.74	1.67
Spain	Wage Structure Survey	1995	118005	8.8	26	7.3	0.071	761	1410	2999	3.94	2.13	1.85
Sweden	Level of Living Surveys	1991	1508	11.8	21.5	4.45	0.070	61	81	127	2.08	1.57	1.33
Switzerland	Labour Force Survey	1995	6334	13.2	19.8	3.6	0.111	23.9	35.9	60.3	2.53	1.68	1.51
UK	Family Expenditures Survey	1995	2183	12.3	22.6	2	0.245	4.1	7.3	13.5	3.33	1.85	1.8
USA	Current Population Survey	1995	42347	12.6	18.5	2.33	0.202	5.5	10	19	3.45	1.82	1.9

See Appendix A for a more detailed characterisation of the data sets.

Results for France and Spain refer to yearly earnings. Hourly wages for France and Spain were computed assuming 1760 h/year. Inequality figures (1, 5, 9) refer to 10th, 50th and 90th percentiles.

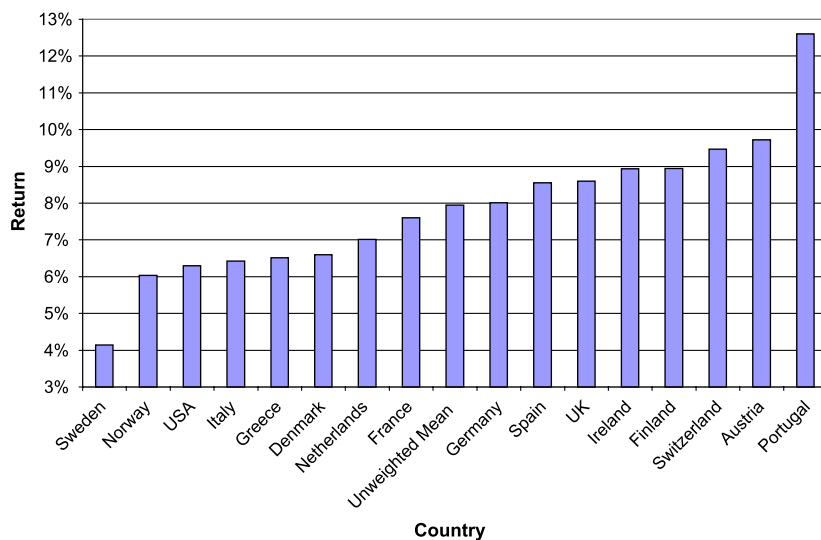


Fig. 1. Returns to education, OLS.

lands, Portugal and Spain). The number of observations used varies reasonably, ranging from fewer than 2000 (Finland, Ireland, Norway and Sweden) to more than 20,000 (Netherlands, Portugal, Spain and the US). All observations refer to full-time male workers only.

All countries draw on gross wages as the measure of earnings. The exceptions are Austria, Greece and Italy, which use net wages, as this was the single type of wage data available. This difference in types of wages may be important, on account of the progressivity of tax systems, and thus trouble the comparability of the results. This will be bore in mind in the following stages of the paper. Hourly wages, the dependent variable, is computed by dividing total wages by total hours in all countries, except for France and Spain. In these two countries, only yearly wages were available, which were then divided by 1760 h.³ Another issue is top coding, which affects the US data set. However, our analysis with a different data set suggests that this is likely not to distort the qualitative results.⁴

³ The sensitivity of the results to the procedure adopted to deal with the lack of hours data was examined by imposing the same number of total hours for all workers in the Portuguese and US data sets. The results, available from the authors, are robust to different procedures.

⁴ We examine the role of top coding by drawing on Portuguese data and artificially censoring the 1% or 5% top of the wage distribution. Again, this matter does not influence the qualitative results. Another issue that may be particularly important in the US is the role of race. However, we extended the main specification with racial dummies (blacks, Hispanics, Asians and native Americans) and found no substantial differences. Both sets of results are also available from the authors upon request.

Table 2
Summary of results

Country	OLS	1st decile	9th decile	Difference
Austria	9.7%	7.2%	12.8%	5.6%
Denmark	6.6%	6.3%	7.1%	0.8%
Finland	8.9%	6.8%	10.1%	3.3%
France	7.6%	5.9%	9.3%	3.4%
Germany	8.0%	7.8%	8.1%	0.3%
Greece	6.5%	7.5%	5.6%	–1.9%
Italy	6.4%	6.7%	7.1%	0.4%
Ireland	8.9%	7.8%	10.4%	2.6%
Netherlands	7.0%	5.3%	8.3%	3.0%
Norway	6.0%	5.5%	7.5%	2.1%
Portugal	12.6%	6.7%	15.6%	8.9%
Spain	8.6%	6.7%	9.1%	2.4%
Sweden	4.1%	2.4%	6.2%	3.8%
Switzerland	9.5%	8.7%	10.6%	1.9%
UK	8.6%	4.9%	9.7%	4.8%
USA	6.3%	3.9%	7.9%	4.0%
Means	7.9%	6.5%	9.1%	2.7%
St. Dev.	2.0%	1.6%	2.6%	2.7%
Coeff. Var.	0.25	0.24	0.29	1.00

Table 1 also presents descriptive statistics from each country's data set. Most countries exhibit levels of average schooling above 10 years, the highest value being that of Switzerland (13.2). The lowest are those of Portugal and Spain, with 6.5 and 8.8 years of average schooling, respectively. Average experience (which corresponds, in all countries, to Mincer experience, age-schooling-6) is generally above 19 and below 22 years. There are a few exceptions, namely Portugal and Spain, with much higher levels (24.5 and 26, respectively), which is not surprising given their low levels of average schooling.

The means and the coefficients of variation of the logarithm of hourly wages are also reported. However, they cannot be compared in a straightforward manner as they are based on different currencies (and slightly different years). The same applies to the hourly wages at the first, fifth and ninth deciles. We used this evidence in the three last columns of Table 1, where we report simple inequality measures, such as the ratios between wages at different deciles.

In the columns headed "Wage ratios", the ratio of the wages at the ninth and first deciles' are generally between 2 and 3. The exceptions are the UK (3.33), US (3.45), Greece (3.62), Spain (3.94), Portugal (4.58) and Ireland (4.74).⁵ Comparing the second and third columns, it can also be seen that in most countries the largest share of the ninth–first deciles inequality is obtained from the top half of the

⁵ With respect to the cases of Austria, Greece and Italy, their measures of inequality should be understood as lower bounds of the true measure, as their wage figures are net of taxes, unlike those of the other countries, and therefore influenced by the progressive nature of the tax systems.

Table 3
Quantile regressions results (coefficients and SEs)

	Austria (1993)		Denmark (1995)		Finland (1993)		France (1993)	
0.1	0.070	0.0034	0.061	0.0026	0.066	0.0067	0.0571	0.00203
0.2	0.075	0.0030	0.062	0.0020	0.083	0.0048	0.0652	0.00197
0.3	0.082	0.0020	0.061	0.0016	0.080	0.0052	0.0682	0.00183
0.4	0.087	0.0028	0.061	0.0017	0.081	0.0042	0.0728	0.00188
0.5	0.091	0.0031	0.061	0.0021	0.088	0.0045	0.0755	0.00174
0.6	0.098	0.0034	0.065	0.0019	0.087	0.0050	0.0809	0.00210
0.7	0.106	0.0042	0.067	0.0024	0.092	0.0048	0.0825	0.00216
0.8	0.113	0.0031	0.069	0.0024	0.092	0.0062	0.0840	0.00290
0.9	0.120	0.0046	0.069	0.0041	0.096	0.0096	0.0890	0.00345
OLS	0.093	0.0021	0.064	0.0018	0.086	0.0042	0.0733	0.00156
	Germany (1995)		Greece (1994)		Ireland (1994)		Italy (1995)	
0.1	0.0748	0.00447	0.073	0.0072	0.075	0.0102	0.065	0.0034
0.2	0.0753	0.00528	0.063	0.0043	0.085	0.0056	0.063	0.0024
0.3	0.0788	0.00324	0.060	0.0041	0.087	0.0048	0.057	0.0021
0.4	0.0785	0.00328	0.059	0.0028	0.089	0.0040	0.057	0.0017
0.5	0.0820	0.00329	0.056	0.0027	0.099	0.0052	0.056	0.0015
0.6	0.0837	0.00318	0.056	0.0028	0.098	0.0055	0.057	0.0019
0.7	0.0865	0.00303	0.055	0.0029	0.100	0.0044	0.061	0.0020
0.8	0.0851	0.00367	0.053	0.0034	0.102	0.0032	0.065	0.0026
0.9	0.0780	0.00437	0.055	0.0047	0.099	0.0049	0.068	0.0033
OLS	0.0803	0.00372	0.063	0.0033	0.086	0.0047	0.062	0.0017
	Norway (1995)		Netherlands (1996)		Portugal (1995)		Spain (1995)	
0.1	0.053	0.0071	0.051	0.0014	0.065	0.0010	0.065	0.0004
0.2	0.048	0.0043	0.054	0.0008	0.083	0.0010	0.076	0.0004
0.3	0.051	0.0042	0.059	0.0008	0.099	0.0009	0.083	0.0004
0.4	0.049	0.0025	0.061	0.0007	0.112	0.0009	0.086	0.0004
0.5	0.056	0.0039	0.063	0.0007	0.122	0.0009	0.087	0.0004
0.6	0.065	0.0044	0.066	0.0008	0.131	0.0011	0.087	0.0004
0.7	0.069	0.0060	0.070	0.0008	0.136	0.0012	0.087	0.0004
0.8	0.070	0.0049	0.074	0.0010	0.140	0.0013	0.087	0.0005
0.9	0.073	0.0080	0.079	0.0013	0.145	0.0017	0.087	0.0006
OLS	0.059	0.0039	0.068	0.0006	0.119	0.0009	0.082	0.0003
	Sweden (1991)		Switzerland (1995)		United kingdom (1995)		USA (1995)	
0.1	0.024	0.0027	0.084	0.0036	0.048	0.0070	0.039	0.0012
0.2	0.028	0.0021	0.084	0.0024	0.056	0.0056	0.050	0.0012
0.3	0.031	0.0022	0.086	0.0022	0.066	0.0053	0.057	0.0011
0.4	0.036	0.0023	0.090	0.0016	0.071	0.0047	0.065	0.0012
0.5	0.043	0.0026	0.092	0.0014	0.070	0.0042	0.068	0.0010
0.6	0.045	0.0025	0.094	0.0018	0.070	0.0036	0.072	0.0009
0.7	0.050	0.0029	0.096	0.0016	0.069	0.0040	0.074	0.0010
0.8	0.055	0.0036	0.100	0.0020	0.075	0.0049	0.075	0.0010
0.9	0.060	0.0044	0.101	0.0026	0.092	0.0060	0.076	0.0015
OLS	0.041	0.0022	0.090	0.0019	0.083	0.0041	0.061	0.0008

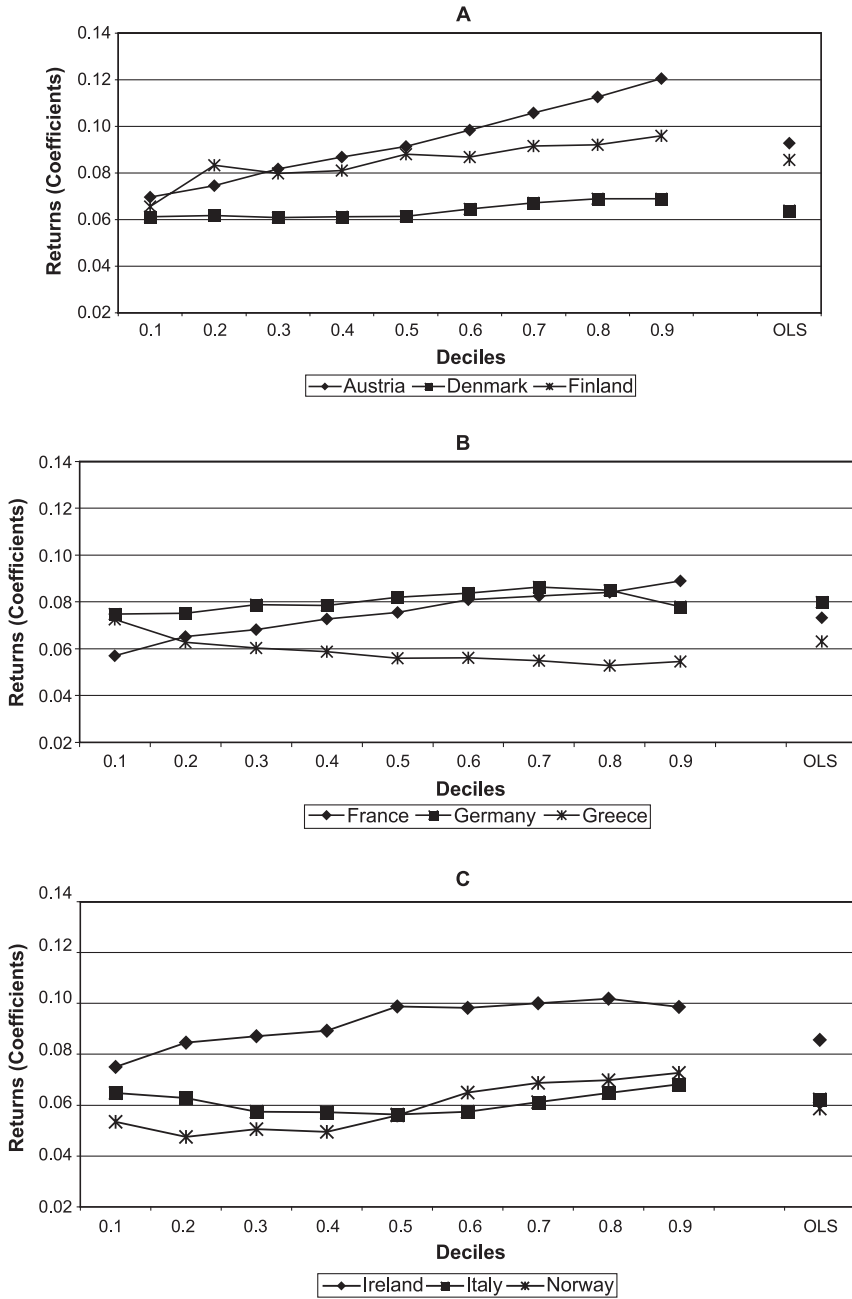


Fig. 2. Returns to education, QR and OLS.

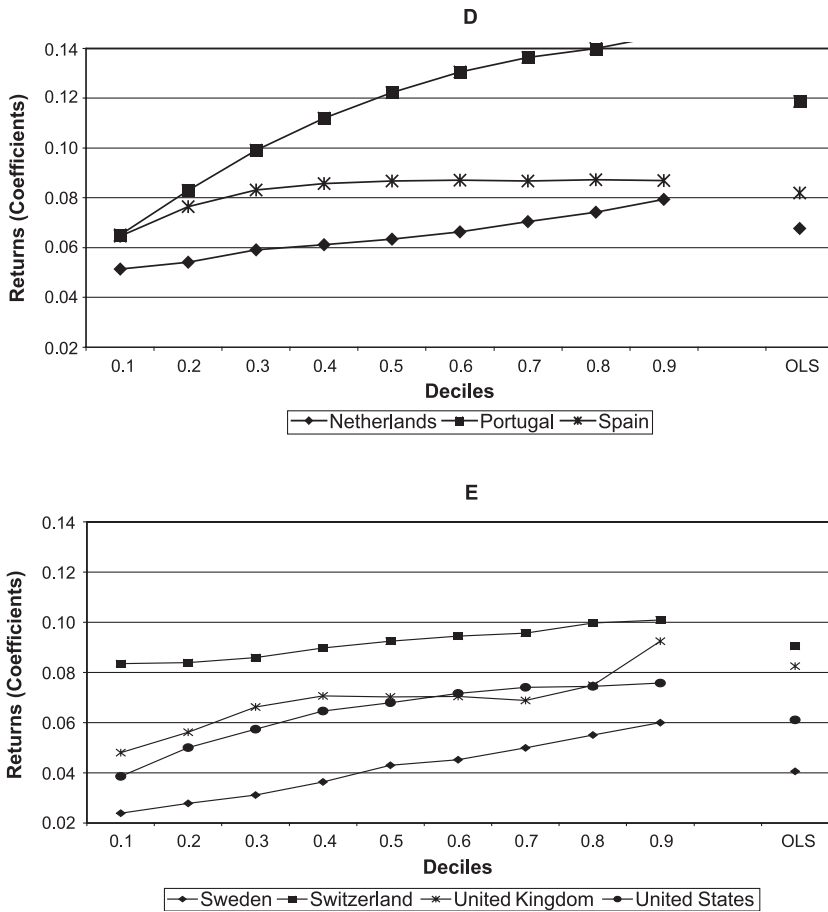


Fig. 2. (continued).

distribution (ninth–fifth deciles), the exceptions being Germany, Greece, Ireland and the US.⁶

4. Empirical results

The empirical results were obtained by regressing the following version of the Mincer (1974) equation, under Becker’s (1975) framework:

$$\log y_i = \alpha_\theta + \beta_\theta \cdot \text{educ}_i + \delta_{\theta 1} \cdot \text{exp}_i + \delta_{\theta 2} \cdot \text{exp}_i^2 + u_i,$$

where $i = 1, \dots, N$ (N being the number of observations for each year), $\theta = 0.1, 0.2, \dots, 0.9$ is the quantile being analysed, y is the hourly wage, educ is the number of schooling

⁶ These results are generally in accordance with those presented at Gottschalk and Smeeding (1997). However, a thorough comparison is impossible as both the time period and the earnings measure covered there are different.

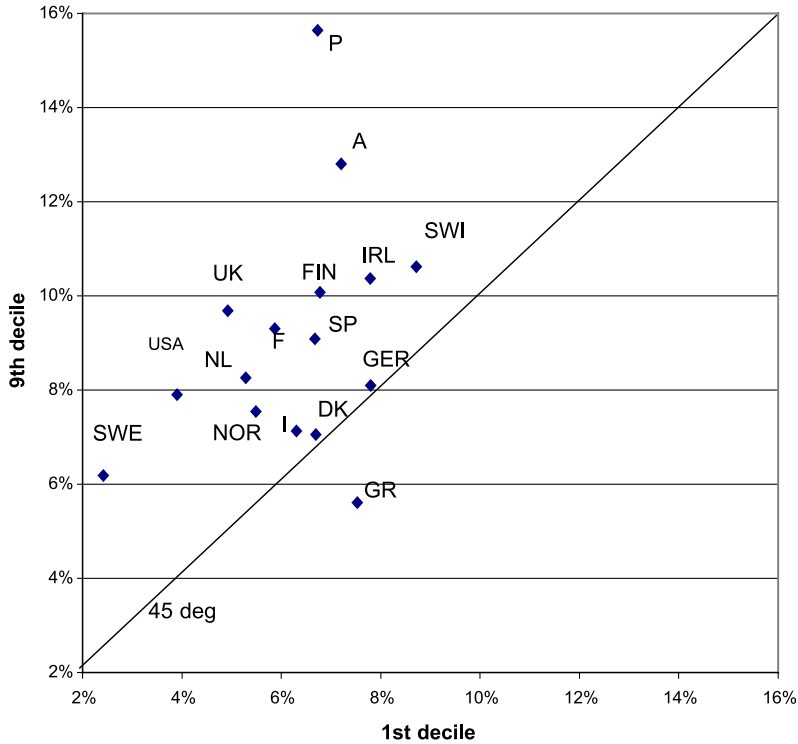


Fig. 3. Returns to education, quantile regression.

years⁷ and exp corresponds to Mincer experience. Only men working full time (35 h or more per week) were considered. The case of women was disregarded on account of the extra complication of potential selectivity biases.⁸

As a benchmark, we first present the results obtained with the traditional OLS method.⁹ In Fig. 1, we rank countries by their OLS return to education. As one can see in Table 2, the mean return is 7.9%, with a standard deviation of 2%. Sweden exhibits the lowest value (4%) whereas Portugal displays the highest (12.6%).

In Table 3, we present the coefficients and associated standard errors for both OLS and quantile regression estimates, for each country. These results are pictured in Fig. 2A–E, where we exhibit both the returns at the mean (OLS) and at different points of the wage distribution (namely the quantiles 0.1,0.2,...,0.9). The stylised fact that comes out from this analysis, which is also the key result of the

⁷ We use information on the highest level achieved. Extra school attainment above the number of school years associated with the degree is thus disregarded.

⁸ For an application of quantile regressions accounting for selectivity issues, see Buchinsky (1998).

⁹ See Asplund and Pereira (1999) and Harmon et al. (2001) for surveys of OLS returns to education across Europe, which were also prepared by the same network involved in the “PuRE” project.

paper, is that returns to education are higher at higher points of the (conditional) wage distribution.¹⁰

The Swedish case is a good example of this general result. Whereas the average return to education is of 4%, the return at the first decile is no greater than 2% and the return at the ninth decile reaches 6%. The more extreme case of this pattern is the Portuguese case. Here an average return of 13% masks a return of only 6% at the first decile and more than 15% at the last decile.¹¹

In the set of 16 countries considered here, only Greece does not follow this pattern. However, the data for this country are based in net wages, which troubles a full comparison with the remaining countries. In particular, progressive taxes are likely to have a strong impact in eroding the returns to education at the top of the distribution than at its bottom. This may explain the Greek results.

A summary graphical description of these results can be found in Fig. 3. Here we only consider the returns to education at the first and ninth deciles for each country. One can see that most countries are lying on the top left part of the figure, which means that their returns at the top of the distribution (proxied by the ninth decile) are higher than those at the bottom (first decile). The clear exception is Greece while Denmark, Germany and Italy are relatively close to the 45° line that separates each situation.¹²

5. Discussion

This paper presents an empirical finding concerning returns to education across the wage distribution. More skilled workers (individuals who receive higher hourly wages conditional on their characteristics) are associated with a stronger education-related earnings increment. We put forward three possible explanations for this result.

A first one lies on over-education. In fact, situations where highly schooled workers take jobs with a low-skill requirement and consequent low pay would be consistent with these results. In this case, the lower earnings of the over-educated will increase the within-skill dispersion of pay by extending the lower tails of the highly educated wage distributions. In the extreme case that the lower tail of the highly skilled workers earnings distribution is fully taken by the over-educated (that is, by highly skilled workers with low-skills jobs), returns to education at the bottom quantiles would be particularly low. One reason why one would expect these returns to be positive is that the over-education

¹⁰ This result is in accordance with previous attempts at estimating wage equations with quantile regressions, such as Machado and Mata (2001), Hartog et al. (2001) and Fersterer and Winter-Ebmer (2003). We have also experimented with other specifications for a subset of countries (Portugal and the US), in particular considering a quartic in experience, following Murphy and Welch (1990). The qualitative results—not shown but available upon request—were unaltered, although we found higher estimates under the more extended specification.

¹¹ Tests on whether these coefficients are significantly different have been performed for most countries and are available upon request from the authors. Except for some countries whose sample sizes are small, these tests support our result of significantly higher coefficients at the upper tail of the distribution.

¹² These data are examined from the point of view of the risk involved in education in Pereira and Martins (2002). In this paper, a significantly positive cross-country correlation between the spread of returns at the top and bottom of the wage distribution and the mean return to education is presented and discussed.

literature (see [Hartog \(2000\)](#) for a survey) finds that, although returns to over-education are lower than the returns to “appropriate” education, they are in any case still positive.

A second alternative or complementary explanation is about ability. In the line of the much-debated results of [Herrnstein and Murray \(1995\)](#),¹³ ability may be the most relevant force in explaining socio-economic success. We take this argument one step further and hypothesise that higher schooling levels may compound the role of ability if there is an interaction between the two variables. This would mean that the role of ability differences within a given schooling level would become increasingly amplified in terms of pay as one considers high schooling levels. While in lower education levels, differences in ability would not be too relevant (and thus the dispersion of earnings would be low), in higher education levels, differences in ability would translate into substantial pay gaps between the high- and low-ability workers.

Finally, differences in school quality or fields of study may be another driving force for the pattern uncovered. Although the Mincer framework considers only differences in school quantity, it may be the case that those individuals that fall into the bottom of their conditional earnings distributions are precisely those who benefited from poorer school quality or who have chosen fields of study with (ex-post) poor returns. These events are likely to jeopardise the earnings potential from these individuals’ human capital, placing them at an increasingly worse position (in relative terms) the higher their educational level. The bottom of the wage distributions would then be over-represented with workers with low-level school quality or who engaged on fields of study that attract scarce interest in the labour market. Moreover, these differences are likely to be more prevalent at higher schooling levels, because those are the stages that exhibit more variety in schooling paths—and possibly also in schooling quality (conditional on the school-leavers at each given school level).

More generally, these explanations can be nested into an interaction between schooling and some factor or set of factors that also impacts upon pay differentials and which are heterogeneously distributed across workers within any given skills level. To the extent that these factors have a stronger influence upon earnings for workers with higher schooling levels, then the empirical stylised fact documented here would be obtained.

A final, important issue concerns the possible impact of endogeneity on the results presented. If high-ability individuals are over-represented in higher-schooling levels (because they face lower marginal costs to schooling, for instance), then not considering ability differences would wrongly lead to an overestimate (underestimate) of the return to schooling at the upper (lower) part of the conditional distributions. This would indeed correspond to the pattern uncovered here.

However, an important recent contribution in this topic, [Arias et al. \(2001\)](#), suggests this is not the case. In this paper, the authors draw on the data from the twins study by [Ashenfelter and Rouse \(1998\)](#) in order to control for ability differences.¹⁴ They find that, although this attenuates the extent of differences in returns to education across the distribution, these returns do still remain significantly different.

¹³ See [Heckman \(1995\)](#) for a critique of the methodology and conclusions of ‘The Bell Curve’.

¹⁴ This procedure—using twins data to identify the returns to education—is not, however, completely satisfactory. For instance, one may be concerned about how exogenous are the differences in schooling between twins which are implicitly used to overcome the endogeneity of the returns to education.

6. Conclusions

In this paper we analyse the dispersion of the returns to education at different Western countries during the mid-1990s, with a view to addressing the link between schooling and within-levels inequality. These issues have not received much attention so far in the literature, which has implicitly assumed that the schooling-related earnings increment is constant across the wage distribution. Notable exceptions are [Card \(1999\)](#) and [Buchinsky \(1994\)](#) who examine these phenomena, both theoretically and empirically—the latter focusing on the US. In particular, [Card \(1994\)](#) asked, ‘Is the labour force reasonably well described by a constant return to education for all workers?’ The evidence we provide suggests that this question should be answered negatively.

In fact, we found a stylised fact over the 16 developed countries we cover: returns to schooling increase over the wage distribution. Or, to put it differently, the earnings increment associated to schooling is higher for those individuals whose unobservable characteristics place them at the top of the conditional wage distribution.

These findings imply that schooling may have a positive impact upon within-group wage inequality, as the spread of returns increases for higher educational levels. One possible explanation for this is over-education (when individuals with higher schooling attainment take jobs requiring lower skills). Another possibility is that there is an interaction between schooling and ability, in which the most able can benefit more from their schooling and the pay gap between the more and less able deepens for higher educational levels. A final explanation may lie on differences in school quality or fields of study. Here, the bottom of the wage distributions would be over-represented with workers with low-level school quality or who engaged on fields of study that attract scarce interest in the labour market. These differences are likely to be more prevalent at higher schooling levels, when there is more variety in schooling paths and schooling quality.

More generally, these explanations can be nested to an interaction between schooling and some factor or set of factors that also impacts upon pay differentials and which is heterogeneously distributed across workers within any given education level. To the extent that these factors have a stronger influence upon earnings for workers with higher schooling levels, then the empirical stylised fact documented here would be obtained.

In any case, regardless of the true source of this result—and this is certainly an area deserving further research—we believe this finding rings some alarm bells at policies designed to cut wage inequality by simply investing in the attainment of higher schooling levels. Even if a given population were made of only highly skilled individuals—which would, on average, command high wages—such population, according to our evidence, would still exhibit considerable levels of pay inequality.

Acknowledgements

This paper was written under the scope of the 15-country ‘PuRE—Public Funding and Private Returns to Education’ European Commission TSER project. The results presented here are based on work done by country teams headed by Mahmood Arai, Rita Asplund, Erling Barth, Giorgio Brunello, Colm Harmon, Joop Odink, Pedro Pereira, José Luis

Raymond, Ali Skalli, Viktor Steiner, Panagiotis Tsakoglou, Ian Walker, Niels Westergaard-Nielsen, Rudolf Winter-Ebmer and Stefan Wolter. We also thank Robin Naylor, participants at several seminars and a referee and the associate editor (D. Black) for comments and suggestions. Financial support from Fundação para a Ciência e a Tecnologia (SFRH/BD/934/2000 and POCTI/ECO/33089/99) and the British Council and research support from the Bank of Portugal are also gratefully accredited. The usual disclaimer applies.

Appendix A. Description of data sets used

A.1. Common features

This section presents the common features across all data sets. Exceptions to this pattern are documented in the following section, which also describes in greater and somewhat varying detail other relevant information about each country's data set.

The information used in all data sets refers to men, aged 15–65, working at least 35 h/week. The dependent variable is the logarithm of gross hourly wages, obtained by dividing weekly, monthly or annual wages by the correspondent number of hours worked. This variable is not censored. Schooling years are obtained from considering the highest grade obtained by the worker and then the number of schooling years typically required for a student to achieve such grade. Experience is Mincer experience (age-education-6).

A.2. Specific Features

Austria: Mikrozensus. A representative 1% household survey, including detailed information about human capital variables. Information on net monthly earnings. All employees (white-collar, blue-collar and civil servants) aged between 15 and 65 years are included in the sample. Apprentices have been eliminated from the analysed population.

Denmark: Longitudinal Labour Market Register. A random 0.5% sample of the adult population, covering the years 1976–1995. All information in LLMR is drawn from administrative registers and is merged by Statistics Denmark. Estimations are based on people aged 16–64.

Finland: Finnish Labour Force Survey. Compiled by Statistics Finland. Representative sample of the whole Finnish population. The sample has traditionally contained some 9000 individuals aged 15–64. Apart from standard individual characteristics, also the information on and income is register based. The rest of the information is self-reported through questionnaires and interviews undertaken by Statistics Finland. The earnings concept refers to the individual's average gross hourly wage as calculated from tax record information on taxable annual earnings and self-reported numbers of months and normal hours worked. The annual earnings comprise all types of compensation, such as overtime and vacation pay, except for fringe benefits.

France: Training and Professional Qualifications. Survey conducted by INSEE, the French national statistics institute. Richest French data set in terms of initial as well as post-school education and their professional outcomes. The household survey called

Enquête Emploi (Employment Survey) is also used. Only FQP gives detailed information on individuals' educational record and family background. In FQP, only gross annual earnings are available.

Germany: German Socio-Economic Panel. Longitudinal household survey conducted on a yearly basis. Detailed information about income, labour market status, education and various other socio-economic variables is collected. The sample was restricted to West German citizens. The self-employed, pensioners, military personnel, people still engaged in education or training were also excluded, as well as foreigners, whose educational background may fundamentally differ from that of native Germans.

Greece: Household Budget Surveys (1993/1994). The only data set available that covers in a consistent way the entire labour force over the last 25 years and contains income information. Carried out by the national statistical service of Greece. It covers the entire non-institutional population of the country and their sampling fraction is 2%. It contains detailed information about consumption expenditures, incomes and socio-economic characteristics of the households and their members. The income component used in the paper is "earnings net of income taxes and social insurance contributions". It includes wages, salaries, overtime payments, bonuses, holiday payments, and related benefits received from the main and secondary employer, normalised on a monthly basis. Further, the surveys report the number of hours normally worked per week. Division of monthly income adjusted on a weekly basis by this figure yields "net hourly earnings". The samples used consist of employees outside the agricultural sector aged 14–64. Thus, self-employed, employers, unpaid family members and apprentices are excluded.

Ireland: Household Survey. Carried out by the Economic and Social Research Institute (ESRI), in 1994. First wave of the Irish part of the European Community Household Panel. Rich survey with regards to labour market experience, including a very comprehensive list of current gross earnings, deductions and net earnings. Includes a measure of the number of schooling years and the highest level of education attained. There is a high level of confidence in the reliability of the data, in terms of how representative it is of the population.

Italy: Survey of Household Income and Wealth. Conducted by the Bank of Italy. Available from 1977 annually and at odd years after 1987. It contains information both on households (family composition) and on individuals. This information includes net yearly earnings, average weekly hours of work and number of months of employment per year. It also contains information on family background (the education, age, occupation and sector of parents). There are no other nationally representative surveys in Italy that cover the same range of information. The sample is restricted to non-agricultural employees aged from 14 to 65.

Netherlands: Structure of Earnings Survey. Conducted by the Dutch Central Bureau of Statistics. Until 1979, the Structure of Earnings Surveys were large cross-sectional employer surveys in which information on gross earnings, educational level, sex, age, and industry of employees was gathered. The 1995 version of the Structure of Earnings Survey was created by combining information at the individual level of three different data sources: the 1995 "Employment and Wages Survey", the 1995 "Insured Persons Register", and the 1994–1996 "Labour Force Surveys".

Norway: Level of Living Surveys. Ongoing project at Statistics Norway, surveying a sample of the Norwegian adult population. Panel survey, adding young persons in every wave. Every LLS contains about 5000 individuals, comprising around 2500 wage observations. The analysis is limited to wage earners, 16–67 years of age. The wage variable is calculated as reported monthly/weekly/hourly gross wage, divided by the number reported weekly hours. The individuals are asked to report their usual level of wages and hours—including usual level of overtime. In the surveys from 1989 and onwards, the educational variable is merged from administrative registers (5-digit code).

Portugal: Personnel Records. Every year since the late 1970s, all firms, either private or public, provide data concerning every employee and also detailed firm-specific information. This data set covers the entire Portuguese labour market (employees) and has a panel structure (at both the employee and employer levels) since the late 1980s.

Spain: Wage Structure Survey 1995. Employer survey of 175,000 wage earners, which contains an important amount of characteristics related to each worker (qualification, tenure, type of contract, type of job, sector, firm size, and so on). All surveys were purged dropping those observations with wages below minimum wage and age below 18 or above 65.

Sweden: Swedish Level of Living Survey, 1991. SLL surveys are the most widely used Swedish data sets for wage equations. SLLS contains about 6000 randomly sampled people between ages 16 and 75 (18–75 for 1991), where the primary sample from 1968 constitutes the base for these surveys: 1974, 1981 and 1991. The non-response rate has increased over time, from 9% in 1968 to roughly 20% in 1991.

Switzerland: Swiss Labour Force Survey. Main data source for labour-market related questions in Switzerland. Produced annually by the Swiss Federal Statistical Office since 1991. The sample of 32,000 in 1995 is representative for the adult population (older than 15 years) permanently living in Switzerland. The data are collected by telephone interviews.

United Kingdom: Family Expenditure Survey. Random sample of approximately 7000 households each year, available since the 1960s. In addition to education and earnings FES contains some information relating to union status and has smoking and other expenditures. Refers to Great Britain (i.e., UK excluding Northern Ireland).

United States: Current Population Survey. Monthly household survey used to address labour-market related issues, e.g., unemployment rates.

References

- Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70, 91–117.
- Arias, O., Hallock, K., Sosa-Escudero, W., 2001. Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics* 26, 7–40.
- Ashenfelter, O., Rouse, C., 1998. Income, schooling and ability: evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113, 253–284.
- Ashenfelter, O., Rouse, C., 2000. Schooling, intelligence and income in America. In: Arrow, K., Bowles, S., Durlauf, S. (Eds.), *Meritocracy and Economic Inequality*. Princeton Univ. Press, Princeton.

- Asplund, R., Pereira, P., 1999. Returns to Human Capital in Europe: A Literature Review. ETLA, Helsinki.
- Becker, G., 1975. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press, Chicago.
- Buchinsky, M., 1994. Changes in the US wage structure, 1963–1987: application of quantile regression. *Econometrica* 62, 405–458.
- Buchinsky, M., 1998. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Econometrics* 13, 1–30.
- Card, D., 1994. Earnings, schooling and ability revisited. NBER Working 4832.
- Card, D., 1999. The causal effect of education on earnings. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labour Economics*, vol. 3. North-Holland, Amsterdam, pp. 1801–1863.
- Fersterer, J., Winter-Ebmer, R., 2003. Are Austrian returns to education falling over time? *Labour Economics* 10, 73–89.
- Gottschalk, P., Smeeding, T., 1997. Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature* 35, 633–687.
- Harmon, C., Walker, I., Westergaard-Nielsen, N., 2001. *Education and Earnings in Europe: A Cross-Country Analysis of Returns to Education*. Edward Elgar, Cheltenham.
- Hartog, J., 2000. Over-education and earnings: where are we, where should we go? *Economics of Education Review* 19, 131–147.
- Hartog, J., Pereira, P., Vieira, J.C., 2001. Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression estimators. *Applied Economics* 33, 1021–1037.
- Heckman, J., 1995. Lessons from the bell curve. *Journal of Political Economy* 103, 1091–1120.
- Herrnstein, R., Murray, C., 1995. *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press, New York.
- Juhn, C., Murphy, K., Pierce, B., 1993. Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101, 3.
- Katz, L., Murphy, K., 1992. Changes in relative wages, 1963–1987: supply and demand factors. *Quarterly Journal of Economics* 107, 35–78.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Machado, J., Mata, J., 2001. Earnings functions in Portugal 1982–1994: evidence from quantile regressions. *Empirical Economics* 26, 115–134.
- Mincer, J., 1974. *Schooling, experience and earnings*. NBER, New York.
- Murphy, K.M., Welch, F., 1990. Empirical age-earnings profiles. *Journal of Labor Economics* 8, 202–229.
- Pereira, P., Martins, P., 2002. Is there a return-risk link in education? *Economics Letters* 75, 31–37.
- Sullivan, D., Smeeding, T., 1997. Educational attainment and earnings inequality in eight nations. *International Journal of Educational Research* 27, 513–525.