46th Congress of the European Regional Science Association
University of Thessaly, Greece, August 30th-September 3rd, 2006

Regional Differences in Returns to Education in Portugal

José António Cabral Vieira
João Pedro Almeida Couto
Maria Teresa Borges Tiago

University of the Azores – Dep. of Economics and Management
Rua Mãe de Deus
9500 Ponta Delgada Portugal

E-mail: josevieira@notes.uac.pt

This paper analyses differences in the return to education in Portugal across regions. For this purpose, we use an extended Mincer-type wage equation. OLS regression results indicate that differences in the rewards to education are substantially different across regions. In particular, they are much higher in Lisbon than in other regions. Since the average level of education in Lisbon is much higher in Lisbon than elsewhere such a differential is attributed to the fact that the demand for educated labour is much higher in Lisbon, likely due to differences in technology. A quantile regression analysis reveals that the return to education is not constant across the whole conditional wage distribution. This is valid for the five regions examined, although once again the impact of education on wages is higher in Lisbon regardless the quantile we examine.

Keywords: regions, returns to education, human capital, Portugal

JEL Code: J31

1. Introduction

The role of regions for wage differentials has been put forward in the literature by several authors (see e.g. Dumond et al., 1999, Duranton and Monastiriotis, 2002 and Bernard et al. (2003). Furthermore it has been widely stated that education contributes to wage formation (see Mincer, 1974, Vieira, 1999, Hartog et al.2001, among many others). In this context, the rate of return to education constitutes a key a key parameter (for a comparison this estimate among countries see Psacharopoulos and Patrinos, 2002). However, it has been shown that the impact of education on wages varies among dimensions such as countries, gender and industries.The main goal of this paper is to provide some evidence on the role of regions to wage determination, as well as to evaluate the size of regional rates of return to education in a small country as Portugal.

This is country for which a few studies have already addressed the effect of regions on wages. For instance, Cardoso (1991) documents the existence of large wage differentials among the Portuguese regions. Vieira (1999) indicates that after controlling for a large number of individual and job attributes employees working in the area of Lisbon and the Tagus Valley earn higher wages than their counterparts in other regions (the lowest wages were paid in the central region of the country). Teulings and Vieira (2004) compare wages in Lisbon and the Tagus Valley with those paid in the rest of the country and argue that higher wages in Lisbon result from differences in the returns to human capital between those two regions. In particular, they argue that equally skilled workers obtain higher returns on human capital due to differences in technology (complexity of the jobs). More recently, Vieira and Madruga (2005) examined low-pay employment incidence and mobility in Portugal and conclude that those working in the region of Lisbon are less likely to be found in the low pay segment and, once in such a situation, are more likely to escape from it.

There is also evidence that the returns to education in Portugal are not constant across regions. For instance, Santos and Vieira (2000) and Vieira et al. (2005) provide evidence that the ‘average’ impact of education on wages varies across regions. In these studies that highest returns are found in the region of Lisbon.

A common feature of most of the aforementioned studies is a high level of aggregation of the regions (in some cases only Lisbon and the Tagus Valley versus the rest of the country), which may to some extent lead to misleading results. Furthermore, most of them use OLS estimators, thus determining the average impact. In this paper, and for empirical purposes, we make use of ordinary least squares (OLS) and quantile regression (QR) estimators. The latest estimator allows us to assess how the effect of education varies across the whole conditional wage distribution. In the OLS perspective, the regression coefficients are assumed constant across the entire conditional wage distribution. However, there is no specific reason to assume in advance such uniformity. The characterisation of the conditional expectation (mean) likely constitutes only a limited aspect of the wage distribution. Indeed, some studies suggest that restricting the analysis to average effects misses important features of the wage structure (e.g. Buchinsky, 1994, Chamberlain, 1994, Machado and Mata, 1997, Fitzenberger and Kurz, 1997).

The paper is organised as follows. Next section describes the data. Section 3 presents the estimation methods. Section 4 includes some theoretical background. Section 5 includes the estimation results. Finally, section 6 concludes and summarizes.

2. Data

The data used here were drawn from *Quadros de Pessoal* (Personnel Records) for 2000. This is a standardised questionnaire which all firms with wage earners must complete every year for the Department of Labour. The data include information on individual workers such as age, tenure with the current firm, the highest completed level of education, and gender. Information is also available on hours of work, firm size, industry affiliation, and regions. Years of education were calculated by attributing the nominal number of completed years in order to complete the reported level in the data. Potential labour market experience was computed as age minus years of education minus six. Hourly wages were calculated as monthly wages divided by the number of hours worked. Civil servants and others serving in the armed forces are not included in the data source. The final sample contains 342 698 non-agricultural, and non-fishermen workers between 16 and 65 years of age. Records with missing values were deleted from the original sample, as were the self-employed, unpaid family workers and apprentices. The data refers only to the mainland.

Some descriptive statistics of the data are included in Table A1 in Appendix. As we can observe, 36% of the individuals in the sample worked in the North, 14% in the Centre, 44% in Lisbon, 3% in Alentejo and 3% in Algarve. Moreover, the highest average level of education is found in Lisbon (8.4 years) and the lowest in the North (6.7 years). The same Table also includes descriptive statistics by region concerning years of labour market experience, years of tenure with the current employer, firm size and the distribution of the workers by gender and industry.

3. Estimation methods

Ordinary least squares is one of the methods used in this analysis. This method allows us to estimate the effect of education on the mean of the conditional wage distribution. However, the impact of education on the mean of that distribution likely describes a partial aspect of the statistical relationship among variables. In such a case, it may be important to examine that relationship at different points of the conditional distribution function. Quantile regression (QR) warrants such an analysis. The QR method was introduced by Koenker and Basset (1978). They define the θth regression quantile as the solution to the problem:

$$\min_{\beta \in R^k} \left[\sum_{(i:y_i \geq x_i'\beta)} \theta \left| y_i - x_i'\beta \right| + \sum_{(i:y_i < x_i'\beta)} (1-\theta) \left| y_i - x_i'\beta \right| \right], \qquad \theta \in (0, 1) \qquad \text{(1a)}$$

This is normally written as:

(1b)

where $\rho_\theta(\varepsilon)$ is the *check function* defined as

$$\rho_{\theta}(\varepsilon) = \begin{cases} \theta\varepsilon & if \quad \varepsilon \geq 0 \\ (\theta - 1)\varepsilon & if \quad \varepsilon < 0 \end{cases}$$

The model specifies the θth-quantile of the conditional distribution of the log-wages, given the covariates *x* as:

$$Q_y(\theta \mid x) = x^{'} \beta_{\theta}, \theta \in (0,1) \qquad (2)$$

By variation of θ, different quantiles can be obtained. The least absolute deviation (LAD) estimator of β is a particular case within this framework. This is obtained by setting θ=0.5 (the median regression). The first quartile is obtained by setting θ=0.25, and so on. As we increase θ from 0 to 1, we trace the entire distribution of *y*, conditional on *x*. This problem does not have an explicit form, but it can be solved by linear programming methods. In this study it is solved by linear programming techniques suggested in Amstrong et al. (1979). In practice, obtaining standard errors for the coefficients in quantile regression is a difficult problem and one for which the literature provides only a sketchy guidance. In the present study we used a bootstrap method with 20 repetitions.

4. Some theoretical background

In order to clarify the importance of the QR technique in a specific context, we present a modified version of the model of optimal schooling choice developed in Card (1994). Assume that an individual chooses education and maximises a utility function of the type:

$$U(w,E) = ln\, w - rE \qquad (3)$$

subject to the individual's opportunity set summarised by *w=g(E),* representing the level of wages (*w*) available at each level of education (*E*). This type of utility function derives naturally by assuming that the individual maximises the discounted present value of wages, discounts the future at a rate *r*, and earns nothing while in school (see Willis, 1986, Card, 1994). The first order condition for optimal education requires that:

$$\frac{g'(E)}{g(E)} = r \qquad (4)$$

In the optimum the marginal rate of return equals the marginal cost of the investment in education.

To make the model empirically operational, we must choose functional forms for the marginal (proportional) benefits and costs of education. For the sake of simplicity, it is assumed that the marginal costs are increasing functions of the amount invested in education, and that the marginal returns do not vary with education (the latter assumption is only a matter of simplicity and can be discarded without changing the main implication). Specifically,

$$\frac{g'(E)}{g(E)} = \beta_i$$

$$r = r_i + kE \qquad (5)$$

Since the individual invests in education until the point where marginal costs equal marginal benefits, his optimal amount of education is given by:

$$E_i^* = \frac{\beta_i - r_i}{k} \qquad (6)$$

Integration of the marginal benefits in (5) leads to a log-linear wage equation for individual *i* of the type:

$$\ln w_i = a_i + \beta_i E_i \qquad (7)$$

Traditionally, variation in ability concerns variation in the intercept of the wage equation. One appealing feature of the model is that variation in ability also concerns the slope. In other words, ability influences the wage-effect of education. If it only influenced the intercept, individuals with higher ability might well invest less in education, since they have a higher opportunity cost of school attendance.

The model identifies two sources of heterogeneity in the population: variation in marginal rates of return to education at each level of schooling (loosely known as differences in ability) and variation in the marginal costs of investment in schooling (loosely known as differences in access to funds or tastes for education). Except under very restricted assumptions, equilibrium in this model implies a non-degenerate distribution of marginal returns to education across the population (Card, 1994). Such a distribution introduces ambiguity into the interpretation of the causal effect of education: in essence, each person has his own causal effect.

This simple model raises an important conceptual question on empirical work. If individuals have different returns to education at the same level of schooling there is no unique causal effect of schooling on wages. The quantile regression technique allows us to shed light onto the issue. The estimation of the effect of education on conditional quantiles permits us to uncover individual heterogeneity in the effect of education on wages. Two examples based on Koenker and Basset (1982), Manski (1988) and Mata and Machado (1995) may help to clarify this point.

Aside from other covariates, consider the following simple wage equation:

$$ln\, w_i = a + \beta E_i + \varepsilon_i \qquad (8)$$

In this equation one can define $a_i = a + \varepsilon_i$ where ε_i are *i.i.d* random terms. Given that specification (8) is correct, heterogeneity among individuals only affects wage levels and therefore concerns the intercept of the wage equation. In such a case,

$$Q_{ln\, w}(\theta|E) = [a + Q_{\varepsilon}(\theta)] + \beta E\ , \qquad \theta \in (0, 1) \qquad (9)$$

Only the intercept differs for different conditional quantiles. The slope - i.e. the marginal effect of *E* - is invariant to the quantile being estimated. The (theoretical) conditional quantile functions form a family of parallel lines. They are parallel to the mean regression line: only the conditional location of the dependent variable changes for different values of θ. In such a case, is no substantial loss of information, with respect to the slope when estimating solely a measure of conditional central tendency such as the mean (estimated by OLS).

However, Koenker and Basset (1982) have warned that when errors are not identically distributed the situation is different. In many applications the conditional quantile function $Q_y(\theta | x)$ probably does not depend on *x* only in location, because the exogenous variables may also influence the scale, tail behaviour, or other characteristics of the conditional distribution of *y* (see Koenker and Basset, 1982, p.49). In such cases, the slope coefficients depend in a non-trivial way on θ and one might expect to find discrepancies in the estimated slope parameters at different quantiles. To clarify the importance of this point consider the (random coefficient) model

$$ln\, w_i = a_i + b_i E_i \qquad (10)$$

where a_i=a + ε_i and b_i=b + ε_i and ε_i is a random variable reflecting individual heterogeneity.

In this case the intercept and the slope coefficient of the theoretical conditional quantile line will vary with the quantile being estimated. If the 'ability' effect concerns only the slope of the wage function (i.e. a_i=a for all individuals), as in most of Card's (1994) set-up, then $Q_{ln\,w}(\theta | E) = a + [b + Q_{\varepsilon}(\theta)]E.$ In any case, b_i= b + ε_i, captures the idea that wages are heterogeneously determined and that the slope coefficient differs in observations with the same observed education. Therefore, there may be information gains from estimating and comparing several conditional location measures for the dependent variable, even after controlling for a large set of observed individual and job characteristics. We will do that for our Portuguese data set, both overall and for several decompositions.

5. Estimation results

This section includes the results of a Mincer-type wage-equation, where the individual's years of education are used as an explanatory variable. Other covariates are a vector of ones, years of tenure with the current firm, a experience and experience squared, firm size, firm age, gender and industries. The dependent variable is the logarithm of hourly wages. The main goal is to estimate the parameter associated with years of education (i.e. the return to education, see Mincer, 1974).

The interpretation of the quantile regression coefficients is conceptually quite analogous to OLS regressions. In OLS case, the regression coefficients measure the influence of the regressor variables on the conditional mean of the dependent variable, whereas in the quantile regression case the coefficients β_θ represent the influence of the regressors on the conditional θ-quantile of the dependent variable.

The marginal effect of a variable on a specific conditional quantile of the dependent variable can be obtained by the corresponding partial derivative. Therefore, 'quantile rates of return to education' are given by:

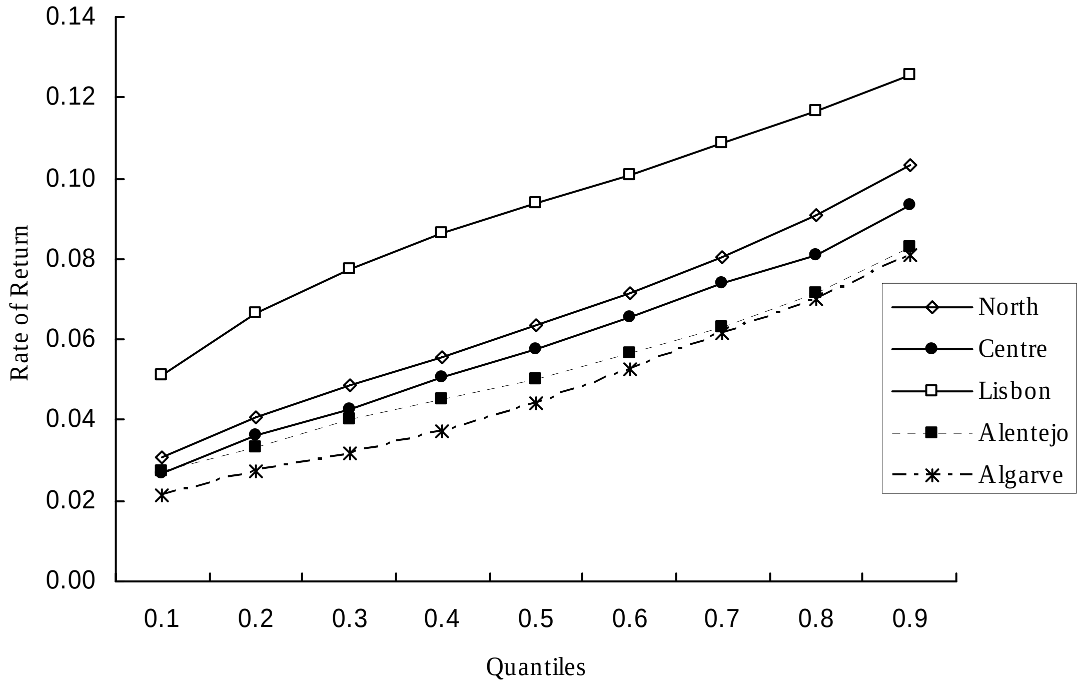
$$r_\theta = \frac{\partial Q_{\ln w}(\theta \mid x)}{\partial E} \qquad (11)$$

The value is multiplied by one hundred to give a percent interpretation.

Nine quantile regressions were computed for each of the three years being examined. Furthermore, the regressions were performed for the full sample, and for two sub-samples of men and women separately. Quantile rates of return to education for the present specification of the wage equation are in Table A2 in the appendix. These are plotted against the quantile numbers in Figure 1.

The effect of education on wages is positive and statistically different from zero at each of the quantiles analysed. This suggests that wages increase throughout the conditional distribution range with education and is valid for the five regions under examination. However, education affects wages differently at different parts of the distribution. It has a larger effect at higher quantiles. his suggests that there is, in all regions, heterogeneity in the returns to education which are larger for individuals at higher (with better-unobserved earning capacity) quantiles of the conditional wage distribution**.** This indicates that modeling on average (i.e. OLS) misses important features of the wage structure, regardless of the region under examination. Finally, the returns to education are always higher in Lisbon than in the other regions, on average and throughout the conditional wage distribution.

Figure 1 - Quantile rates of return to education by region



6. Conclusions

This paper was an attempt to provide a comprehensive picture of the returns to education by region in Portugal. For this purpose, we used two estimation methods. The results indicate that there is much heterogeneity in the returns to education. The results also indicate that the effect of education on wages is not equal across the conditional wage distribution, regardless of the region. Returns are higher for individuals with higher positions in the conditional distribution. Apparently, the labour force is not reasonably described in any region by a constant (average) effect of education on wages. These results indicate that modelling on average (i.e. OLS) misses important features of the wage structure.

Finally, the returns to education are higher in Lisbon than in the other regions. Since the (average) supply of educated labour is higher in Lisbon, we may argue such as Teulings and Vieira (2004) that higher returns in this region are eventually due a higher demand associated to differences in technology (complexity of the jobs).

References

Amstrong, R.; Frome, E. and Kung, D. (1979) Algorithm 79-01: a revised simplex algorithm for the absolute curve fitting problem, in Communications in Statistics, Simulation, and Computation B8(2), Marcel Dekker, New York.

Bernard, A., Redding, S. Schott, P. And Simpson, H. (2003) "Regional wage variation and industry location", The Institute for Fiscal Studies, mimeo.

Buchinsky, M. (1994) Changes in the U.S. wage structure 1963-1987: application of quantile regression, *Econometrica*, 62, 405-458.

Card, D. (1994) Earnings, schooling and ability revisited, Working Paper No. 4832, National Bureau of Economic Research, Cambridge.

Cardoso, A. (1991) *Desigualdades salariais em Portugal: uma perspective regional 1982-1989*, Masters Dissertation, Lisboa, Instituto Superior de Economia e Gestão.

Chamberlain, G. (1994) Quantile regression, censoring and the structure of wages, in Proceedings of the Sixth World Congress of the Econometrics Society, Barcelona, Spain (Eds.) C. Sims and J.J. Laffort, Cambridge University Press, New York, 171-209.

Dumond, M., Barry, H. and Macpherson, D. (1999) "Wage differentials across labour markets: does cost of living matter?", *Economic Inquiry*, 37, 577-598.

Duranton, G. and Monastiriotis, V. (2002) "Mind the gaps: the evolution of regional earnings inequality in the U.K.", *Journal of Regional Science*, 42, 219-256.

Fitzenberger, B. and Kurz, C. (1997) New insights on earnings trends across skill groups and industries in Germany, Diskussionspapiers 38, Center for International Labor Economics, Universitat Konstanz.

Hartog, J., Pereira, P. and Vieira, J. (2001) "Changing Returns to Education in Portugal during the 1980s and Early 1980s: OLS and Quantile Regression Analysis", *Applied Economics*, 33, 1021-1037.

Koenker, R. and Basset, G. (1982) Robust tests for heteroscedasticity based on regression quantiles, Econometrica, 50, 43-61.

Koenker, R. and Basset, G. (1978) Regression quantiles, Econometrica, 46, 33-50.

Machado, J. A. and Mata, J. (1997) Earning functions in Portugal 1982-1994: evidence from quantile regressions, Paper presented at the workshop The Portuguese Labour Market in International Perspective, Lisbon, July 18-19.

Manski, C. (1988) Analog Estimation Methods in Econometrics, Chapman & Hall, London.

Mincer, J. (1974) *Schooling, experience and earnings*, New Tork: National Bureau for Economic Research.

Mata, J. and Machado, J. (1995) Firm start-up size: a conditional quantile approach, European Economic Review, 40, 1305-23.

Psacharopoulos, G. and Patrinos,, H. (2002) "Returns to Investment in Education: A Further Update".World Bank Policy Research Working Paper No. 2881.

Santos, C. and Vieira, J. (2000) "Análise da Decomposição Inter-insular das Diferenças Salariais na Região Autónoma dos Açores", in *Perspectivas de Desenvolvimento para as Regiões Marítimas*, Actas do VII Encontro Nacional da APDR, Volume 1, pp 101-107.

Teulings, C. and Vieira, J. (2004) "Urban versus rural return to human capital in Portugal: a cookbook recipe for applying assignment models", *Labour*, 18, 265-291.

Vieira, J. (1999) *The evolution of wage structures in Portugal, 1982-1992*, Ph.D. Dissertation, Amsterdam, Tinbergen Institute Research Series, 197.

Vieira, J. and Madruga, P. (2005) "Regions and low-wage mobility in Portugal", *Regional and Sectoral Economic Studies*, 4, 41-52.

Vieira, J., Couto, J. and Tiago, M. (2005) "Inter-regional wage dispersion in Portugal", Institute for the Study of Labor (IZA), Bona, Discussion Paper Nº 1664.

Appendix

Table A1 – Sample descriptive statistics

	North		Centre		Lisbon		Alentejo		Algarve	
	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.	mean	std. dev.
log hourly wage	6.480	0.480	6.494	0.455	6.853	0.621	6.525	0.468	6.553	0.455
years of education	6.680	3.401	6.840	3.443	8.374	3.981	7.015	3.490	7.244	3.544
years of tenure with the current employer	7.304	8.395	6.667	7.874	8.034	9.291	6.052	7.677	4.714	6.624
Years o labour market experience	22.69	12.02	23.40	12.57	23.50	13.00	23.633	13.133	23.61	13.37
log of firm size	3.883	1.957	3.549	1.671	4.826	2.521	3.105	1.735	2.995	1.528
male	0.579	0.494	0.582	0.493	0.594	0.491	0.599	0.490	0.535	0.499
manufacturing	0.311	0.463	0.170	0.376	0.052	0.222	0.139	0.346	0.044	0.205
wood, cork, paper and chemistry	0.139	0.346	0.254	0.435	0.114	0.318	0.146	0.353	0.052	0.222
electronics and transp. equipments	0.074	0.262	0.072	0.258	0.044	0.205	0.039	0.193	0.005	0.072
electricity, gas, water and construction	0.123	0.328	0.123	0.329	0.108	0.310	0.135	0.342	0.129	0.335
retail and wholesale, hotels and restaurants	0.209	0.407	0.228	0.420	0.293	0.455	0.341	0.474	0.532	0.499
Banking, financing and transportation	0.053	0.224	0.035	0.185	0.180	0.384	0.022	0.145	0.048	0.213
Real state and services provided to firms	0.035	0.183	0.032	0.177	0.118	0.323	0.050	0.218	0.091	0.288
Health, education and social services	0.039	0.194	0.070	0.254	0.059	0.236	0.107	0.309	0.064	0.245
Social, personal and domestic services	0.017	0.183	0.015	0.171	0.033	0.164	0.021	0.287	0.035	0.271
# of observations	124023		47721		150856		9658		10440	

Table A2 - Rates of returns to education: OLS and quantile regression estimators

	North		Centre		Lisbon		Alentejo		Algarve	
	coeff.	std. error	coeff.	std. error	coeff.	std. error	coeff.	std. error	coeff.	std. error
OLS	0.0764	0.0004	0.0666	0.0006	0.0981	0.0004	0.0627	0.0013	0.0532	0.0014
Quantile:										
.10	0.0307	0.0003	0.0268	0.0005	0.0511	0.0003	0.0272	0.0014	0.0212	0.0011
.20	0.0406	0.0003	0.0364	0.0005	0.0667	0.0003	0.0334	0.0011	0.0273	0.0011
.30	0.0486	0.0003	0.0429	0.0004	0.0773	0.0003	0.0404	0.0015	0.0320	0.0011
.40	0.0558	0.0003	0.0507	0.0006	0.0861	0.0003	0.0453	0.0013	0.0372	0.0011
.50	0.0637	0.0003	0.0578	0.0006	0.0937	0.0004	0.0503	0.0012	0.0439	0.0015
.60	0.0717	0.0003	0.0654	0.0007	0.1009	0.0004	0.0564	0.0016	0.0526	0.0016
.70	0.0804	0.0005	0.0737	0.0008	0.1087	0.0005	0.0628	0.0019	0.0615	0.0020
.80	0.0909	0.0006	0.0811	0.0011	0.1167	0.0007	0.0715	0.0023	0.0702	0.0028
.90	0.1034	0.0010	0.0933	0.0016	0.1258	0.0009	0.0831	0.0032	0.0811	0.0042