
Returns to education and wage equations

PEDRO TELHADO PEREIRA and PEDRO SILVA MARTINS*†

Universidade da Madeira, Campus da Penteada, Madeira, Portugal and

†University of St Andrews and Department of Economics, University of Warwick, Coventry, CV4 7AL, UK

The paper shows why considering a number of education-dependent covariates in a wage equation decreases the coefficient of education in that equation. This result is illustrated empirically with a meta-analysis for Portugal. The education coefficient decreases when covariates are used that can be considered post-education decisions; on the other hand, it is independent of sample size, tenure and whether hourly or monthly wages are used. These results support the use of a simple specification of the Mincer equation for the study of the total returns to education.

I. INTRODUCTION

Although the debate on the causal link between education and productivity is still ongoing, policymakers are already drawing on results from Mincer equations to support their decisions regarding the optimal private cost of (higher) education. However, it is well known that Mincer equations are sensitive to the inclusion of extra covariates, and this fact has brought some confusion to the public debate. This paper sheds some light on this matter, by addressing questions such as ‘Why does the inclusion of industries, for instance, in the wage equation decrease the return to education by so much?’

Education is one of the many investment decisions motivated by the fact that the investment yields a choice that one would not otherwise have. Part of the return to the investment is to be found in the set of options that emerges. For instance, when an individual decides upon the level of education to be attained, it is believed that such academic qualification will lead to a better-paid job. That qualification will also extend the number of options in other matters, as well, such as the sector and/or specific firm where the individual will be employed. Part of the individual’s return to education will thus be the return to subsequent choices—choices that are available only after qualification is obtained.

In this respect, an examination of the literature reveals two distinct main lines of research: the ‘economics of education’ branch, which focuses on the total return to education, and the ‘labour economics’ branch, which seeks to explain wage differentials among individuals.

We see these two lines of research as complements. In this paper, we give a simple example to illustrate their differences (Section II), explain the relationship between them (Sections III and IV), and test our findings by means of a meta-analysis using data for Portugal. Section V concludes.

II. THE PARADIGM OF THE TWO ISLANDS

Let us imagine that there are two islands, one (I_1) with a productivity per capita of P and the other (I_2) with a productivity per capita of Q , with $P < Q$. The inhabitants of the islands live an eternal life (all are born at the same moment) and maximize the present value of their production.

The inhabitants of I_1 can move to I_2 if they spend one period learning how to swim. The only cost is the product they forgo during that period. The discount rate is uniformly distributed between (r_1, r_2) ; $f(r) = 1/(r_2 - r_1)$.

*Corresponding Author. E-mail: p.martins@warwick.ac.uk

The decision of learning to swim is made by comparing

$$(E_1) \quad \sum_{i=0}^{\infty} \frac{P}{(1+r)^i}$$

and

$$(E_2) \quad \sum_{i=1}^{\infty} \frac{Q}{(1+r)^i}$$

There is a value of r , r_c , such that $E_1 = E_2$. Let us assume that $r_1 < r_c < r_2$. If $r < r_c$ then $E_1 < E_2$ and the individual decides to learn to swim and thereafter swims from I_1 to I_2 . If $r > r_c$ then $E_1 > E_2$ and the individual does not learn to swim and stays on I_1 .

Therefore, there will be S_1 proportion of swimmers, where

$$S_1 = \int_{r_1}^{r_c} f(r) dr$$

We end up with three groups of individuals: (1) living on island 1, (2) living on island 2 and born there, and (3) living on island 2 and born on island 1.

If we want to explain the differences in productivity we can do so by examining the place where the person is living (the wage equation approach). This is what the ‘labour economists’ do. But suppose that we are interested in studying the ‘returns’ to education (learning to swim). We must then look at the persons born on island 1 and see the differentials in their productivities as the return we are looking for. This is what ‘education economists’ look for. We could never find this return if we considered the place of residence as one of the explanatory variables of the productivity differential, the reason being that this covariate is a result of learning to swim for people born on island 1.

This simple example illustrates the difference between wage equations, where education is one of the explanatory variables, and returns to education, where all the indirect effects should be accounted for.

III. THEORETICAL BACKGROUND

Becker (1962) and Mincer (1974), in their breakthrough contributions for the economics of education,¹ advanced a very appealing equation specification, given that it harmoniously matches inductive and deductive evidence. Drawing on moderately weak assumptions, they prove that running the following specification, $\log y = \alpha + \beta educ +$

$\delta_1 exp + \delta_2 exp^2$, with cross-section data,² β represents the rate of return to education.³

The precise definition of the variables to be used remains unclear, however. Income, for instance, might be either net or gross, hourly, weekly, monthly or yearly. Furthermore, the above-mentioned equation specification has been extended on a number of occasions by the addition of several different controls seeking to explain wages differences. The efficiency wages hypothesis (see, *inter alia*, Krueger and Summers, 1988) warrants the inclusion of sectors of activity, firm size and firm age. The existence of ‘rents’ and trade unions or agency models (Freeman and Medoff, 1986; Pencavel, 1991; Hart and Holmstrom, 1987) justify the inclusion of the bargaining regimes. Internal wage structures (Lazear, 1998) justify the inclusion of seniority (tenure). Wage equations have thus been estimated using all or subsets of these variables.

It is well known that model mis-specification biases all coefficients estimated. However, if we want to see the full impact of education on wages, we have to consider the impact of education in other explanatory variables (covariates) and their effects on wages, therefore we can conclude that:

- (1) To obtain the full effect of education on wages, one should be careful not to include in the wage equation covariates whose value can depend on education. In the extreme case we should only regress the $\ln(\text{wage})$ in education.
- (2) If we include in the regression covariates that depend on education, then the coefficient of education decreases (at least in the expected value of the estimator).

IV. APPLIED DISCUSSION

Instead of engaging in data mining, e.g. testing different wage equation specifications and examining their implications in the education coefficient, we use a meta-analysis of results from other studies.

In a nutshell, a meta-analysis is a regression that takes as dependent variable the outcomes from different studies that focus on the same topic and employ the same general methodology. The regressors describe the characteristics (in terms of equation specification, in sample size, in year of estimation, and so on) underlying those different results and/or studies. A meta-analysis is therefore a useful tool for summarizing several results on a given topic, allowing a researcher to have a global and quantifiable view on the

¹ See also Card (1999) for a thorough survey of the returns to education literature and Bjorklund and Kjellstrom (2000) for criticism of the Mincer specification.

² y represents income, $educ$ stands for the total years of education and exp represents labour-market experience.

³ Of course, the exogeneity of education is an important assumption. We do not consider this in this paper.

link between the structure of a research project and its results. In the present study, we examine the influence of covariates in the return to education.

We use results for Portugal, given the large number of studies undertaken for this country. From the first known paper published on the issue—Psacharopoulos (1981)—until the recent ‘PuRE, Public funding and private returns to education’ project⁴ (Pereira and Martins, 2001), many estimates have been produced. Moreover, we draw on a stepwise estimation procedure to choose the variables to include in the meta-analysis regression. We also correct the estimated standard errors, since the use of an estimated value as the dependent variable makes the model heteroscedastic. Finally, only coefficients obtained with ordinary least squares are considered, as the evidence from other estimation methods (for instance, instrumental variables) is very scarce in Portugal. We use results for males as a way of avoiding sample selection issues.

As stated previously, we are not interested in examining the particular effect of a certain covariate in the coefficient of education, but rather in studying the impact of including covariates upon the education coefficient.

Data description

A meta-analysis uses two kinds of data, which we label here as foreground and background data. The former is directly used information, which includes the coefficients of education that were obtained in different studies, and the presumably relevant characteristics of those studies. By such characteristics, we mean the regressors used, sample size, and so on. Background data, on the other hand, is simply the primary sources (data sets) used for computing the returns to education (or better, a coefficient to education). In this section, we describe both types of data.

Table 1 presents the different papers/projects from which we extracted the information we used. These papers cover the main available results on returns to education in Portugal.⁵

Table 2 presents descriptive statistics of the data we gathered from the above-mentioned studies. Coefficient of education, *coeff* (in percentage terms⁶) range between 3.2 and 11.5 and average 9. Bearing in mind that these estimates were obtained by OLS methods, and comparing them to similar results from other European countries, these are substantially high estimates.

Table 1. Summary of papers/projects for Portugal

Author	Year of data used
Hartog <i>et al.</i> (1999)	1982, 86, 92
Kiker <i>et al.</i> (1997)	1991
Kiker and Santos (1991)	1985
Machado and Mata (2001)	1982, 94
Martins (1991)	1977
Psacharopoulos (1981)	1977
Pure project (1998–2000)	1982–1995
Vieira <i>et al.</i> (1997)	1982, 86, 92
Vieira (1999)	1986, 92

Table 2. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min.	Max.
<i>coeff</i>	86	0.090	0.014	0.032	0.115
<i>explvar</i>	86	4.023	7.754	0	35
<i>ssize</i>	86	20333.650	9847.250	392	42347
<i>priv</i>	86	0.163	0.371	0	1
<i>public</i>	86	0.163	0.371	0	1
<i>privpub</i>	86	0.326	0.471	0	1
<i>interac</i>	86	0.047	0.212	0	1
<i>age</i>	86	0.058	0.235	0	1
<i>monthly</i>	86	0.523	0.502	0	1
<i>hours</i>	86	0.105	0.308	0	1
<i>pure</i>	86	0.791	0.409	0	1
<i>year95</i>	86	-7.395	4.633	-18	0
<i>barg</i>	86	0.186	0.391	0	1
<i>regs</i>	86	0.233	0.425	0	1
<i>fage</i>	86	0.035	0.185	0	1
<i>fsize</i>	86	0.198	0.401	0	1
<i>fowner</i>	86	0.151	0.360	0	1
<i>ten</i>	86	0.302	0.462	0	1
<i>sector</i>	86	0.058	0.235	0	1
<i>regs1</i>	86	0.023	0.152	0	1
<i>ten1</i>	86	0.093	0.292	0	1
<i>sum3</i>	86	0.012	0.108	0	1
<i>sum4</i>	86	0.012	0.108	0	1
<i>sum5</i>	86	0.151	0.360	0	1
<i>sum6</i>	86	0.035	0.185	0	1

The number of explanatory variables, *explvar*, in the background regressions (see Table 3) corresponds to the number of regressors used besides education and a constant. Two is the most common number (66%) since in most occasions only experience and experience squared were added in the regression. The maximum was 37 in the sample used.

⁴This is a policy-focused project, tackling the relationship between education systems differentiation and labour-market outcomes. It draws together research teams from 15 European countries. The authors are members of the Portuguese team. More information at www.etla.fi/PURE.

⁵We restricted our attention to those results that use a single regressor for education. Therefore, those regressions that instead used dummies (each standing for a different educational level) were ignored.

⁶This means that the regression coefficients were multiplied by 100.

Table 3. *Distribution of explanatory variables*

Var	Freq.	Per cent	Cum.
2	57	66.28	66.28
3	1	1.16	67.44
4	2	2.33	69.77
5	4	4.65	74.42
6	2	2.33	76.74
7	2	2.33	79.07
15	13	15.12	94.19
26	3	3.49	97.67
37	2	2.33	100.00

Table 4. *Distribution of years of estimates*

Year	Freq.	Per cent	Cum.
1977	4	4.65	4.65
1982	6	6.98	11.63
1983	5	5.81	17.44
1984	5	5.81	23.26
1985	13	15.12	38.37
1986	8	9.30	47.67
1987	5	5.81	53.49
1988	5	5.81	59.30
1989	5	5.81	65.12
1991	7	8.14	73.26
1992	7	8.14	81.40
1993	5	5.81	87.21
1994	5	5.81	93.02
1995	6	6.98	100.00

Table 5. *Distribution of sample sizes*

Size	Freq.	Per cent	Cum.
0–9,999	16	18.60	18.60
10–19,999	10	11.63	30.23
20–29,999	53	61.63	91.86
30–39,999	4	4.65	96.51
> 40,000	3	3.49	100.00
Total	86	100.00	

In the remainder of this section, we present the variables that we think can influence the results obtained in the different regressions. First we start by the year of estimation, *year* (Table 4). Regressions are distributed in a balanced way, 1985 being the year that was submitted to more regressions. Attention should be drawn to the long period (19 years) which was covered by the different papers surveyed.⁷

⁷ This variable was recoded, so that the intercept could have a more interesting interpretation: 0 replaced 1995, –1 was used instead of 1994, and so on.

⁸ As we did not reject that the coefficients of these two variables were different, we created a variable *privpub* as the sum of the two.

⁹ This is due to the fact that we made a correction in the number of years of education of one of the grades. When following up the same worker in different years we noticed that for a certain technological degree the majority of people had nine years of education and not eleven as previously considered.

Sample size, *ssize* (Table 5) is the number of observations used in each regression. Even if the smallest number is 392, 98% of the samples had more than 1000 observations and more than 80% were larger than 10,000, the largest having more than 40,000.

The remaining controls had a qualitative nature and appear in a dummy format equal to one if the control is accounted for in the wage equation; for instance, if there was control for sectors, the dummy *sector* appears with the value 1. We can see in Table 2 that almost 6% of the regressions controlled for it.

The base of comparison is a regression of a Mincer equation with hourly wages as the dependent variable and education in years and computed experience (age minus years of education minus 6) and its squared as explanatory variables, for the year of 1995.

The characteristics include: (1) public/private, *priv*, *public*⁸ (whether samples are for the individual who works in a public or in a private firm); (2) monthly wages with control for hours, *hours*; (3) monthly wages without control for hours, *monthly*; (4) *age* instead of computed experience (the type of labour-market participation control used); (5) *PURE* (if the estimates were produced in our own research);⁹ (6) *interact* (if the regressions took into account the possibility of interaction between education and experience).

Other controls included in the studies are: regions (*regs*), bargaining regime (*barg*), firm age (*fage*), firm size (*fsize*), firm ownership (*fowner*), tenure (*ten*) and *sector*. We tried to use each of these variables independently but the results were very unstable due to the high correlation between them, as authors tend to maintain a certain specification in all the regressions they present.

As a solution to the problem of correlation we created six variables, as there are studies where: (1) only one of them appears (*regs1* if regions and *ten1* if tenure), (2) three of them appear (*sum3* *regs*, *ten* and *sector*), (3) four of them appear (*sum4* *regs*, *fsize*, *ten*, *sector*); (4) five of them appear (*sum5* *regs*, *barg*, *fsize*, *fowner* and *ten*) and (5) (*sum6* *regs*, *barg*, *fage*, *fsize*, *ten*, *sector*). As mentioned above we are interested in the sign of the effect of the use of the covariates in the coefficient of education and not in the effect of a particular control.

We now direct our attention to the so-called background data. Studies of the returns to education in Portugal draw overwhelmingly on a comprehensive Portuguese dataset: *Quadros de Pessoal* (QP, henceforth). Every year, Portuguese firms have to submit to the Ministry of Employment

Table 6. Regression results

Number of obs = 86 $F(7, 76) = 1096.92$ Prob > $F = 0.0000$ $R\text{-squared} = 0.8561$				
	Robust			
Coeff	Coef.	Std. Err.	t	$p > t $
sum4	-0.0565	0.0025	-22.249	0.000
sum3	-0.0298	0.0025	-11.664	0.000
sum6	-0.0319	0.0025	-12.588	0.000
age	-0.0213	0.0020	-10.337	0.000
sum5	-0.0190	0.0013	-14.064	0.000
year95	0.0008	0.0001	5.917	0.000
privpub	-0.0106	0.0014	-7.259	0.000
regs1	0.0073	0.0031	2.353	0.021
PURE	0.0105	0.0025	4.097	0.000
_cons	0.0973	0.0027	35.583	0.000

Regression with robust standard errors.

information on both firm characteristics and those of their employees. This information is very rich, providing well over 25 relevant regressors.

Another attractive feature of this data set is its very large size, which obviously ensures more precise estimates. Researchers work, in most cases, with 2.5% samples (some 50,000 workers per year) but this figure has risen to 25% or even 100% (approximately 500,000 or 2 million observations). The main drawbacks of this data set lie in the lack of household information and the non-representative nature of workers, given that public servants, self-employed and people outside the labour market are not represented.

Results

We used a stepwise procedure to select the model, as we were interested in examining what variables significantly affected the coefficient of education. We tested for the equality of the coefficients of public and private ownership and could not reject the null. The forward and backward procedure gave the same result.

The fact of considering the interaction of education with experience, monthly wages with or without control for hours, sample size and tenure (as the only additional explanatory variable) do not seem to influence significantly the coefficient of education. This is what we expected as the value of these variables are not dependent of choices due to education.

Interpretation of the results

Constant. Our regression produced an intercept of 9.7%, which can be roughly interpreted as the value one would get with 1995 data considering all the other relevant variables which appear in the table equal to 0 and independent of the value of the variables which were dropped from the estimation.¹⁰

The 95% interval for the constant is from 0.0919 to 0.0102, more or less 0.006 around the mean. All the other coefficients are (in absolute terms) higher than this value with the exception of the coefficient of year95.

Sample year. There is a positive relationship between the year of the data that was used and the size of the coefficient of education. In fact, all the studies that have undertaken an analysis of returns to education in different years in Portugal have come up with a clearly increasing trend. According to our results, returns increase by an average of 0.0009 each year, increasing almost 1% per decade. We have also tested for possible nonlinearity in the evolution of returns by adding a squared year term to the equation. We did not reject the hypothesis that the coefficient is equal to zero at any reasonable significance level, so we retained the linear specification.

This result contrasts with the general idea that returns to education fall along with a country's development (on account of less-binding liquidity constraints or more generous public support schemes, for instance). In fact, this would increase the supply of skilled individuals, thus decreasing the reward of such qualification in the labour market. Of course, demand-side considerations must also

¹⁰ This value is very similar to the one obtained in the sample used in the PURE study (9.6) for the 1995 estimates with the standard Mincer equation.

be taken into account: as a country develops, one would expect that higher qualifications become more rewarded by businesses. Taking both explanations together, it would ensue that the price of labour skills would depend on the relative shifts of both the demand and the supply curves, and such a price could either fall or rise.

This scenario of having both the demand and the supply curves of skilled workers shifting outward fits the recent Portuguese economic history. On the one hand, we witness a somewhat pronounced movement of workers from labour-intensive industries to capital-intensive sectors. On the other hand, there has been a significant increase in the human capital endowments of the Portuguese workers, albeit the (still) very low average educational attainment (less than seven years of schooling). It might thus be the case that the increase in demand for skilled workers has been relatively more powerful than the corresponding increase in supply.

Age. As expected, this variable appears with a negative sign, as people do get older as they go on studying. The value of the coefficient is almost symmetric of the one obtained for experience in the Mincer equation (Pereira and Lima, 1999).¹¹

Privpub. Using samples that use only public firms or only private firms has a negative effect in the coefficient of education in both cases. Further studies are needed, but a possible explanation is that the intercept (the constant) is different for both samples and compensating for different work conditions and risk of unemployment.

PURE. The positive coefficient comes as no surprise as we consider nine years of education for a group of workers for which the previous studies considered 11 years. This was only possible because we could construct panel data and after 1994 the technological degree was divided, so we could know who had 9 and 11 years of education. We saw that the large majority had only 9 years of education, so we considered 9 years for this type of education instead of the 11 years, as in other studies.

Regions. The positive coefficient of regions appears to be puzzling if we assume that the choice of the region is only based in terms of best paid jobs and people choose the region to live only after they finish their education. As there are costs of moving from region to region and family ties, the sign of the coefficient can somehow explain this lack of flexibility as well as the fact that not only monetary factors influence people's decisions.

Other variables. All the other coefficients are negative, which supports the conclusion at we arrived in the previous section. They range in absolute value from 0.019 to 0.056. The highest value is obtained when the sector of activity is among the controls used in the wage equation and can

reduce the coefficient of education to half of its size. This leads us to question if the choice of the sector should not be considered as part of the returns to education, and what the nature of this education/sector link is.

V. CONCLUSIONS

The use of the Mincer equation in its simpler form seems to give an approximate value for the total return to education. If more covariates are used in this equation and these covariates are choice variables that depend on education, then it is shown that the coefficient of the education should fall.

This result is supported by the meta-analysis we performed using data for Portugal. The coefficient decreases with all combinations of variables used and can drop to half of its size, especially when the sector of activity is one of the covariates used. The education-related choice of sector is an aspect that should reflect itself in over-education in the better paying sectors.

The increase of the return to education when regions is used as one of the covariates needs further research, as it seems to show that in the Portuguese case the mobility due to job opportunities is rare. Sample size, the use of monthly wages instead of hourly wages, the interaction between education and experience and tenure do not seem to influence the coefficient, which shows its robustness to sample size, specification of the simple Mincer equation and variables that are independent of education.

We also find that the return to education in Portugal in 1995 is around 9.7% and that it increased by about 1% over ten years. This increase in the returns has been going on at the same time that there has been a large increase in the average education of new workers in the labour market, perhaps indicating a larger increase in the demand for skills.

There are a number of future research directions. As pointed out above the influence of education in the choice of sectors and other decisions taken after school should be taken into account when one studies the full benefits brought by education to an individual. As in our simple model of two islands, returns to education and changes in productivity can be very distinct realities. Both are worth studying but one should distinguish between them.

REFERENCES

- Becker, G. (1962) Investment in human capital: a theoretical analysis, *Journal of Political Economy*, **70**, 9–49.
 Bjorklund, A. and Kjellstrom, C. (2002) Estimating the return to investments in education: How useful is the standard Mincer equation? *Economics of Education Review*, **21**, 195–210.

¹¹ It should equal if the specification was linear in experience instead of squared.

- Card, D. (1999) The causal effect of education on earnings, in *Handbook of Labour Economics*, (eds) O. Ashenfelter and D. Card, Vol. 3, North-Holland, Amsterdam.
- Freeman, R. B. and Medoff, J. L. (1986) The two faces of unionism, in *Readings in Labor Economics and Labor Relations*, (eds) L. G. Reynolds, S. H. Masters and C. H. Moser, 4th edn, Prentice Hall, Englewood Cliffs, NJ.
- Hart, O. and Holmstrom, B. (1987) A theory of contracts, in *Advances in Economic Theory, Fifth World Congress*, (ed) T. F. Bewley, Cambridge University Press, Cambridge.
- Hartog, J., Pereira, P. T. and Vieira, J. C. (1999) Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression estimators, *Applied Economics*, **23**, 1021–37.
- Kiker, B., Santos, M. and Oliveira, M. (1997) Overeducation and undereducation: evidence for Portugal, *Economics of Education Review*, **16**, 111–25.
- Kiker, B. and Santos, M. (1991) Human capital and earnings in Portugal, *Economics of Education Review*, **10**, 187–203.
- Krueger, A. and Summers, L. (1988) Efficient wages and the inter-industry wage structure, *Econometrica*, **56**, 259–93.
- Lazear, E. P. (1998) *Personnel Economics for Managers*, John Wiley and Sons, New York.
- Machado, J. and Mata, J. (2001) Earnings functions in Portugal, 1982–1994: evidence from quantile regressions, *Empirical Economics*, **26**(1), 115–34.
- Martins, A. (1991) Human capital earnings functions: the Portuguese case. Faculdade de Ciências Económicas e Empresariais da Universidade Católica Portuguesa, Working Paper.
- Mincer, J. (1974) *Schooling, Experience and Earnings*, National Bureau of Economic Research, New York.
- Pencavel, J. (1991) Higher education, productivity, and earnings: a review, *Journal of Economic Education*, **22**, 331–59.
- Pereira, P. T. and Lima, F. (1999) Wages and human capital: evidence from the Portuguese data, in *Returns to Human Capital in Europe – A Literature Review*, (eds) R. Asplund and P. T. Pereira, ETLA, Helsinki.
- Pereira, P. T. and Martins, P. (2001) Portugal, in *Education and Earnings in Europe – a Cross Country Analysis of Returns to Education*, (eds) C. Harmon, I. Walker and N. W. Nielsen, Edward Elgar, Cheltenham.
- Psacharopoulos, G. (1981) Education and the structure of earnings in Portugal, *De Economist*, **129**, 532–45.
- Sappington, D. (1991) Incentives in principal-agent relationships, *Journal of Economic Perspectives*, **5**, 45–66.
- Vieira, J. C., Hartog, J. and Pereira, P. T. (1997) A look at changes in the Portuguese wage structure and job level allocation during the 1980s and early 1990s. Tinbergen Institute, Discussion Paper TI 97-008/3.
- Vieira, J. C. (1999) Returns to education: a comparison of alternative estimators, *Labour Economics*, **6**, 535–41.

AUTHOR QUERIES

JOURNAL ID: RAEF-100960

**QUERY
NUMBER**

QUERY

1

Citation for Table 6 is not provided. Please check the placement