The Total Impact of Schooling on Within-Groups Wage Inequality in Portugal

Corrado Andini *

ABSTRACT

Using Portuguese data from the 2001 wave of the European Community Household Panel, we analyze to what extent the endogeneity of schooling affects the estimation of the total impact of schooling on within-groups wage inequality by means of quantile-regression techniques. We conclude that the standard techniques assuming schooling-exogeneity may underestimate the total impact of schooling.

Keywords: Endogeneity, Quantile Regressions, Schooling, Wage Inequality.
JEL Classification: I21, J31, C29.

* Assistant Professor, University of Madeira, Department of Management and Economics, Funchal, Portugal. Email: andini@uma.pt, Tel.: +351 291 70 50 53, Fax: +351 291 70 50 40,

Financial support by the European Commission (EDWIN Project, HPSE-CT-2002-00108) is gratefully acknowledged. For their useful comments, I am indebted to Santiago Budría, Günther Lang, Joop Hartog, Pedro Telhado Pereira, Rudolf Winter-Ebmer, and Robert Wright. The usual disclaimer applies.

I. INTRODUCTION

There is substantial empirical evidence supporting the argument that schooling has a positive impact on within-groups wage inequality in Portugal. Most of this evidence is based on the estimation of several types of wage equations using the standard quantile-regression techniques due to Koenker and Bassett (1978). As a matter of example, Hartog *et al.* (2001), Machado and Mata (2001) as well as Martins and Pereira (2004) use different econometric specifications of a wage equation and all find that the return to schooling is increasing over the conditional earnings distribution. Particularly, the return at the ninth decile of the wage distribution seems to be significantly higher than the return at the first decile.

A common feature of the quantile-regression studies on schooling and wage inequality in Portugal is that no one deals with the endogeneity of schooling which is, typically, a relevant issue in studies focusing on the average impact of schooling on earnings. On this point, indeed, there is an open discussion in the literature about whether or not the ordinary-least-squares estimator is biased when the underlying empirical model disregards individual abilities and measurement errors. Some authors argue that the bias does exist, other authors argue that the bias in one direction associated with disregarding individual abilities is somehow compensated by the bias in the opposite direction associated with disregarding measurement errors.

Whether or not a kind of compensation really happens for the average return to schooling, it is less likely that such compensation happens for the return to schooling *at each* decile (or even quantile) of the wage distribution. In other words, the standard quantile-regression techniques applied to a wage equation, without controlling for individual abilities and measurement errors, might provide inconsistent estimates of the return to schooling *at some* deciles. Further, if inconsistency happens at the extreme deciles of the wage distribution (for instance the first and the ninth), then the estimated impact of schooling on within-groups earnings dispersion may be misleading.

Another feature of existing studies is that no one deals with the *total* impact of schooling on within-groups wage inequality, meaning that previously-estimated wage equations do not exclude schooling-dependent covariates, such as industry dummies[1], labour-market experience, and so on (see Pereira and Martins, 2004)[2].

This paper tackles a controversial issue. On the one hand, if the potential correlation between errors and schooling is limited by the insertion in the wage equation of a large set of control-variables, standard quantile-regression techniques are likely to consistently estimate the coefficient of schooling over the conditional wage distribution but unlikely to recover the total returns to schooling due to the likely presence of schooling-dependent covariates among control-variables. On the other hand, if the set of control-variables excludes schooling-dependent covariates and only includes strictly exogenous regressors, then standard techniques may poorly estimate the total impact of schooling on within-groups wage inequality due to the likely correlation between errors and schooling. We aim at providing some empirical evidence on this issue.

II. EMPIRICAL MODELS AND RESULTS

Let us consider the following simple linear model:

[1] Jobs in some industries may require more years of schooling than jobs in other industries.

[2] Pereira and Martins (2004) properly argue that in order "to obtain the full impact 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 one should only regress the ln(wage) in education." (p. 526). See also Andini (2006).

(LM) $Y = \beta X + \xi$

The ordinary-least-squares estimator (say OLS) and standard quantile-regression estimator due to Koenker and Bassett (1978, say KB) are based, respectively, on the following hypotheses (among others):

(H-OLS) $E(\xi | X) = 0$

(H-KB) $\mathrm{Quant}(\xi_\theta | X) = 0 \quad$ for each θ,

where θ is an indicator of the distribution quantile. If X is endogenous, then both hypotheses are violated and both estimators, OLS and KB, are inconsistent.

This paper compares estimates of β based on the hypothesis that X is exogenous (OLS and KB) with estimates of β allowing for X to be endogenous. Specifically, we use the well known two-stage-least-squares approach (2SLS) and the two-stage-instrumental-variable-quantile-regression approach on the lines of Arias *et al.* (2001, say AHS). Because of our interest in the *total* return to schooling, we follow Angrist and Krueger (1991) in the choice of exogenous (years of birth) and instrumental variables (quarters of birth). The empirical analysis is therefore based on the following two empirical models and associated estimators:

(KB) $\ln w_i = \kappa_\theta + \beta_\theta s_i + \sum_j \delta_{\theta j} yb_{ji} + \varepsilon_{\theta i}$

with $\mathrm{Quant}(\varepsilon_{\theta i} | s_i, yb_{ji}) = 0$ for each θ

(AHS) $$\ln w_i = \kappa_\theta + \beta_\theta \hat{s}_i + \sum_j \delta_{\theta j} yb_{ji} + \omega_{\theta i}$$

$$s_i = \nu + \sum_j \alpha_j yb_{ji} + \sum_k \phi_k qb_{ki} + \eta_i$$

$$\hat{s}_i = E(s_i | yb_{ji}, qb_{ki})$$

$$\text{with } \operatorname{Quant}(\omega_{\theta i} | \hat{s}_i, yb_{ji}) = 0 \text{ for each } \theta$$

Our notation is as follows. Letter i indicates the i-th individual in the sample, $\ln w$ stands for the logarithm of hourly gross wage, s measures years of schooling, yb is an indicator-variable for year of birth with j going from 1937 to 1984 (year 1936 is the excluded category), qb is an indicator-variable for quarter of birth with k going from 1 to 3 (quarter 4 is the excluded category).

Data are extracted from the last available wave of the European Community Household Panel, that of 2001. Summary sample statistics are reported in Table 1.

The first-stage regression is based on ordinary-least-squares[3] (see Table 2). As well known, quarters of birth must be considered weak instruments if the F-test of excluded instruments does not reject the null, like in our case. It is also known, however, that passing the F-test should not be intended as a strict requirement due to the limitations of the test itself (low power, see Cruz and Moreira, 2005). This is particularly true when

[3] This explains why we consider declared schooling-years instead of education levels. Specifically, the use of education levels would have implied the estimation of a first-stage linear-probability model and therefore the implementation of maximum-likelihood techniques, which are not allowed by assumption (see model AHS). Alternatively, we might have transformed individual levels of education into equivalent (successfully completed) schooling years, but such procedure would have implied an arbitrary treatment of unobservable measurement errors.

the model specification is not rejected (see the Sargan test). Nevertheless, the presence of weak instruments should not be ignored as it increases the likelihood that an estimator assuming the exogeneity of explanatory variables provides the same answer as an estimator keeping potential endogeneity into account (see Bound *et al.* 1995, Table 3). Weak instruments, indeed, imply two main consequences: first, instrumental-variable estimates (may be inconsistent as well, and) converge to standard estimates; second, standard errors get larger.

Figures 1 and 2 plot the estimation results[4] reported in Table 3. The ordinary-least-squares estimator seems to underestimate the average effect of schooling on wages, but our evidence is somehow mixed because the coefficient obtained using the two-stage-least-squares estimator is not statistically significant due to an expected loss of efficiency.

The main empirical result of this paper is related to Figure 3, that jointly plots the conditional returns associated with model KB and the boundaries of the confidence region associated with model AHS. Specifically, the AHS method seems to suggest a steeper pattern of the conditional return to schooling over the wage distribution than the one predicted by the KB method, *despite* the presence of weak instruments.

III. CONCLUSION

Overall, we provide additional empirical support to the argument that schooling has a positive impact on within-groups earnings dispersion in Portugal. Indeed the total impact of schooling, measured as difference between the return at the ninth and the return at the first decile, is likely to range between the 4% estimated by the standard quantile-regression techniques and the 26% estimated using instrumental-variables.

[4] These graphs are obtained using STATA and the GRQREG module due to Azevedo (2004).

REFERENCES

Andini, C. (2006) Returns to Education and Wage Equations: a Dynamic Approach, *Applied Economics Letters*, forthcoming.

Angrist, J.D. and Krueger, A.B. (1991) Does Compulsory Schooling Attendance Affect Schooling and Earnings, *Quarterly Journal of Economics*, 106(4), 979-1014.

Arias, O., Hallock, K.F. and Sosa-Escudero, W. (2001) Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data, *Empirical Economics*, 26(1), 7-40.

Azevedo, J.P. (2004) GRQREG: Stata Module to Graph the Coefficients of a Quantile Regression, Statistical Software Components, Department of Economics¸ Boston College.

Bound, J., Jaeger, D.A. and Baker, R.M. (1995) Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak, *Journal of the American Statistical Association*, 90(430), 443-450.

Cruz, L.M. and Moreira, M.J. (2005) On the Validity of Econometric Techniques with Weak Instruments, *Journal of Human Resources*, 40(2), 393-409.

Hartog, J., Pereira, P.T. and Vieira, J.A.C. (2001) Changing Returns to Education in Portugal during the 1980s and Early 1990s: OLS and Quantile Regression Estimators, *Applied Economics*, 33(8), 1021-1037.

Koenker, R. and Bassett, G. (1978) Regression Quantiles, *Econometrica*, 46(1), 33-50.

Machado, J.A.F. and Mata, L. (2001) Earnings Functions in Portugal 1982-1994: Evidence from Quantile Regressions, *Empirical Economics*, 26(1), 115-134.

Martins, P.S. and Pereira, P.T. (2004) Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries, *Labour Economics*, 11(3), 355-371.

Pereira, P.T. and Martins, P.S. (2004) Returns to Education and Wage Equations, *Applied Economics*, 36(6), 525-531.

Table 1. Summary sample statistics

Variable	Obs.	Mean	S.E.	Min	Max
Logarithm of hourly gross wage	1782	6.55	0.47	3.32	8.49
Schooling years	1782	8.80	3.91	3.00	27.0
Year of birth	1782	1965	11.5	1936	1984
Quarter of birth	1782	2.45	1.11	1.00	4.00

Table 2. First-stage regression of schooling years

	Coeff.	Robust S.E.	t	P-value
Quarter of birth				
1	-0.5303	0.2733	-1.94	0.053
2	-0.5162	0.2780	-1.86	0.064
3	-0.3004	0.2675	-1.12	0.262
Year of birth				
1937	-0.1968	0.4193	-0.47	0.639
1938	0.4486	0.8273	0.54	0.588
1939	0.1668	1.1782	0.14	0.887
1940	-0.2851	0.3714	-0.77	0.443
1941	1.2821	1.2007	1.07	0.286
1942	0.0151	0.7509	0.02	0.984
1943	2.6175	1.3599	1.92	0.054
1944	0.4641	1.3862	0.33	0.738
1945	3.5350	1.1231	3.15	0.002
1946	1.7825	0.9022	1.98	0.048
1947	0.1080	0.5157	0.21	0.834
1948	2.4706	1.2484	1.98	0.048
1949	1.3233	0.7644	1.73	0.084
1950	2.5171	0.9047	2.78	0.005
1951	1.4911	0.6433	2.32	0.021
1952	1.2574	0.6183	2.03	0.042
1953	1.2702	0.4896	2.59	0.010
1954	1.5278	0.8376	1.82	0.068
1955	0.5447	0.4065	1.34	0.180
1956	1.6778	0.6129	2.74	0.006
1957	1.6236	0.6541	2.48	0.013
1958	1.4677	0.5944	2.47	0.014
1959	0.6664	0.4798	1.39	0.165
1960	1.1161	0.4322	2.58	0.010
1961	1.6299	0.6509	2.50	0.012
1962	2.3032	0.5585	4.12	0.000
1963	2.2418	0.5484	4.09	0.000
1964	1.4711	0.5304	2.77	0.006
1965	3.2459	0.7877	4.12	0.000
1966	3.1548	0.8257	3.82	0.000
1967	3.8398	0.9662	3.97	0.000
1968	3.0704	0.6384	4.81	0.000
1969	3.1249	0.5733	5.45	0.000
1970	3.7298	0.6285	5.93	0.000
1971	3.8072	0.4555	8.36	0.000
1972	3.9569	0.5066	7.81	0.000
1973	3.1319	0.4842	6.47	0.000
1974	4.3609	0.5063	8.61	0.000
1975	4.4149	0.5147	8.58	0.000
1976	3.6441	0.4565	7.98	0.000
1977	3.9069	0.4371	8.94	0.000
1978	3.5783	0.4024	8.89	0.000
1979	3.8873	0.4110	9.46	0.000
1980	3.9421	0.3377	11.67	0.000
1981	4.0952	0.2941	13.92	0.000
1982	2.7029	0.4154	6.51	0.000
1983	2.6506	0.3425	7.74	0.000
1984	2.6250	0.4438	5.91	0.000
Constant	6.4083	0.2554	25.09	0.000
Sargan test of over-identifying restrictions				0.1351
F-test of excluded instruments				0.1972

Table 3. Conditional returns to schooling

Quantile	KB OLS	AHS 2SLS
5	0.0211 (0.0052)	0.0550 (0.0828)
10	0.0298 (0.0033)	0.0000 (0.0782)
15	0.0352 (0.0025)	0.0054 (0.0536)
20	0.0368 (0.0019)	0.0028 (0.0376)
25	0.0396 (0.0017)	0.0000 (0.0155)
30	0.0427 (0.0023)	0.0000 (0.0151)
35	0.0497 (0.0018)	0.0000 (0.0108)
40	0.0518 (0.0020)	0.0214 (0.0526)
45	0.0537 (0.0018)	0.0329 (0.0329)
50	0.0570 (0.0032)	0.0221 (0.0398)
55	0.0577 (0.0026)	0.0549 (0.0613)
60	0.0614 (0.0031)	0.0977 (0.0473)
65	0.0631 (0.0030)	0.1312 (0.0489)
70	0.0658 (0.0029)	0.1541 (0.0567)
75	0.0698 (0.0036)	0.2195 (0.0397)
80	0.0723 (0.0038)	0.2473 (0.0729)
85	0.0716 (0.0031)	0.3029 (0.0795)
90	0.0717 (0.0042)	0.2624 (0.1123)
95	0.0706 (0.0062)	0.1808 (0.1073)
Mean	0.0577 (0.0029)	0.0751 (0.0534)

Standard errors in parentheses

Figure 1. Conditional returns to schooling and confidence regions using KB and OLS

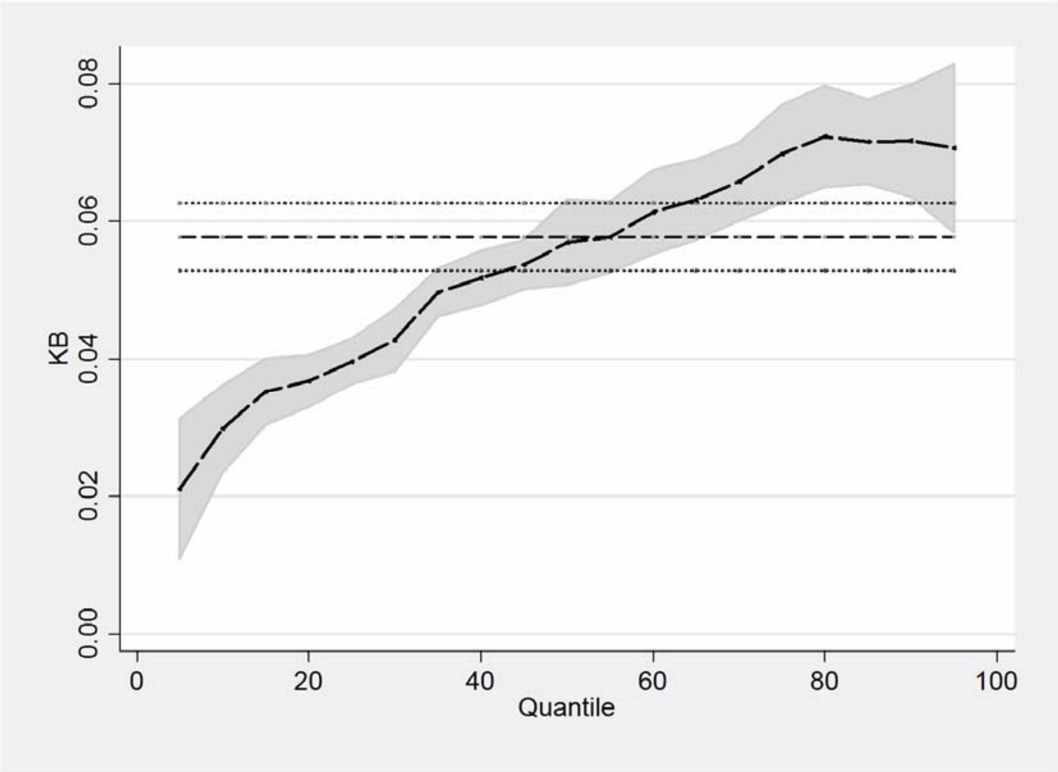

Figure 2. Conditional returns to schooling and confidence regions using AHS and 2SLS

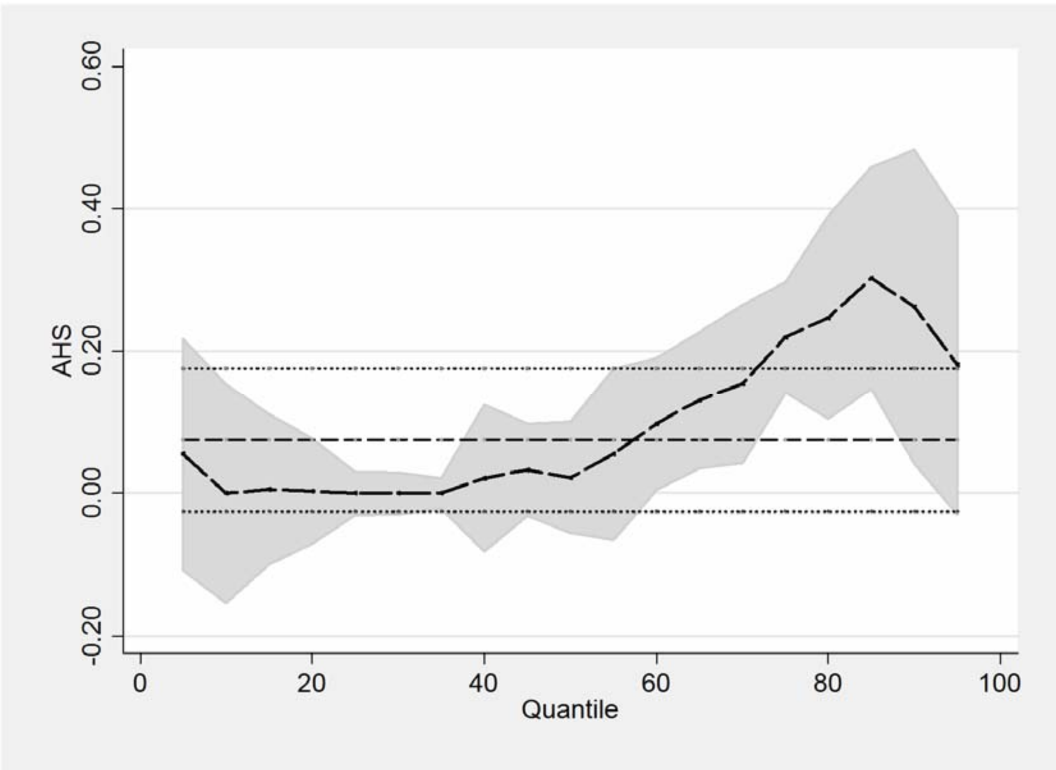

Figure 3. Conditional returns to schooling using KB and confidence region using AHS