Inter-industry Wage Premia in Portugal: Evidence from EU-SILC data

Biagetti, Marco and Scicchitano, Sergio

University of Rome "La Sapienza", Ministry of Economic Development, Italy

8 December 2009

Inter-industry Wage Premia in Portugal: Evidence from EU-SILC data

- **Marco Biagetti[a] and Sergio Scicchitano[a,b *]** -

[a] Ministry of Economic Development, Department for the Development and the Economic Cohesion, via Sicilia 162, 00187, Rome, Italy.

[b] University "La Sapienza", Faculty of Economics, Department of Public Economics, Via del Castro Laurenziano 9, 00161, Rome, Italy.

This version December 8th 2009

Abstract

In this paper we investigate whether inequality in the inter-industry wage premia may be explained by unobserved differences in workers' educational skills. We use the 2007 EU-SILC data set for Portugal, a nation which can be considered a case-study, due to its high inter-industry wage dispersion. Applying both OLS and quantile regression techniques, our results suggest that this unobserved heterogeneity is not a relevant matter in the wage premia determination. We thus corroborate the previous empirical contribution to *Economic Letters* performed by Martins (2004).

Keywords: Returns to education; inter-industry wage inequality; Quantile regression; Portugal
JEL Classification: C14, I21, J31.

* Corresponding author. E-mail: sergio.scicchitano@uniroma1.it, Tel: +39-06-47619923, fax +39-06- 42086946.

1. Introduction[1]

There is a wide consensus that the inter-industry wage inequality across countries remains a relevant topic. As a matter of fact, since the seminal paper by Krueger and Summers (1988) some empirical analysis have explored possible explanations. Particularly, unobserved quality differences across workers have also been called to explain the behavior of income inequality among industries, in different countries (e.g., Gibbons and Katz, 1992).

In this context, Portugal is a chiefly suitable country, because its inter-industry wage dispersion is found to be high when compared with other European countries (Hartog et al. 2001).

In particular, in a contribution to *Economics Letters,* Martins (2004, M henceforth), by applying a Quantile Regression (QR) approach to a 1995 Portuguese data set, pointed out that unobservable differences across workers are not a critical element in determining industry wage premia in that country. He tested the relevance of the unobserved worker quality hypothesis by comparing the differences in the returns across high and low-wage industries and evaluating the correlation between OLS returns and both QR coefficients and inter-quantile differences. The intuition is that for the unobserved ability explanation to be relevant, high wage industries should also show a large difference in returns between the top of the wage distribution, where high ability workers are expected to be found, and its bottom, where are likely to be low ability employees. In this paper we apply a similar approach, using the last 2007 wave of the European Union Statistics on Income and Living Conditions inquiry (EU-SILC), the new European homogenized panel survey. More specifically, our primary purpose is to explore the potential for EU-SILC data to shed some light on the inter-industry inequality in Portugal's wage premia.

[1] The views expressed in this article are those of the authors and, in particular, do not necessarily reflect those of the Ministry of Economic Development. The usual disclaimer applies.

To address these issues we apply the QR semi-parametric approach which is more interesting, as well as more suitable, for it allows us to get a more precise picture of the dynamics of the dependent variable at different points of the distribution, rather than at the conditional mean. We deeply examine QR and OLS, in order to provide a cross-industries comparable view.

The paper supports the idea that in Portugal there is not only a relevant cross-industries dispersion in returns to education, but also a high heterogeneity in wage premia at different points of the wage distribution, which OLS modelling of conditional average of a dependent variable completely fails to account for. In particular, 12 years later the data used in M's analysis and through a different data-set, we confirm that unobservable quality difference across workers cannot be called to spell out Portuguese inequality in the inter-industry wage premia.

This empirical paper is organized as follows. The next section describes the data. Section 3 illustrates our econometric specification. Section 4 reports the results as well as a robustness check. Section 5 display the robustness check, while in the last section we present our main conclusions.

2. Data

Data are collected from the 2007 EU-SILC survey – version 1 of March 2009 - containing information for 24 European countries amongst which Portugal. It is the new homogenized panel survey that has replaced European Community Household Panel (ECHP).

Our analysis focuses on Portuguese full-time male workers aged between 25 and 65: women have been disregarded on account of potential selectivity biases. Younger males have been also dropped because they are still in the almost exclusively educational period of their life, i.e. they are very much likely enrolled in a secondary or tertiary course than performing a work activity whatsoever. In the 2007

EU-SILC wave 4,665 Portuguese men are interviewed: yet, our dependent variable - the hourly (logarithmic) gross wage – is available for 1,735 full-time Portuguese male-workers aged between 25 and 65 years. These constitute our reference sample. Our regressors are schooling years, which has been built up following the usual framework[2], the number of years spent in paid work and its squared: the second is regarded as being a proxy for individual experience while the third takes account of possible non linearities. Sectors surveyed are those considered in EU-SILC, i.e. the 12 macro-sectors obtained by joining several NACE 1.1 classification codes (see table 1 for further details).

[Table 1 here]

3. Econometric specification: OLS versus quantile regression

We assume our OLS specification to have the following simple form:

$$\ln w_i = \alpha_i + \beta_i S + \gamma_i X + \delta_i X^2 + \varepsilon_i \qquad (1)$$

As is well known, OLS implicitly assume that the impact of the regressors along the conditional distribution of the response variable are irrelevant. But as is already equally known covariates may influence it

[2] That usual framework refers to making use of the highest ISCED level of education attained by a male worker, and for each level assigning the legal minimum number of years typically required to achieve it: more precisely, those who reached only an ISCED 1 grade have been given 5 years of schooling; 8 years of school have been assigned to those with an ISCED 2 grade; 13 years to those with an ISCED 3; 14 to people who attained an ISCED 4 grade and 18 years to those who reached an ISCED 5

on its whole shape. This case can be studied by performing a quantile regression (QR), which has the following functional form (Koenker & Basset, 1978):

$$Quant_\theta \left\{ \begin{array}{l} \sum_{i:\ln w_i \geq x_i\beta} \theta \mid \ln w_i - \left(\alpha_\theta + \beta_\theta \mathrm{S}_i + \gamma_\theta X_i + \delta_\theta X_i^2\right) \mid + \\ + \sum_{i:\ln w_i \leq x_i\beta} (1-\theta) \mid \ln w_i - \left(\alpha_\theta + \beta_\theta \mathrm{S}_i + \gamma_\theta X_i + \delta_\theta X_i^2\right) \mid \end{array} \right\} \quad (2)$$

The equation (2) is normally written as:

$$\min_{\beta \in R^k} \sum_i \rho_\theta \left(\ln w_i - \alpha_\theta - \beta_\theta \mathrm{S}_i - \gamma_\theta X_i - \delta_\theta X_i^2 \right) \quad (3)$$

where $\varrho_\theta(z) = \theta z$ if $z \geq 0$ or $\varrho_\theta(z) = (\theta - 1)z$ if z<0.

This problem is solved using linear programming methods. Standard errors for the vector of coefficients are obtainable by using a bootstrap procedure.

4. Results

In table 2 we show OLS returns to education as well as conditional returns at 5 representative quantiles in the 12 Portuguese industries. Differences between percentiles of the wage distribution computed for 4 different extremes taken by twos (θ90-θ10 and θ75-θ25) are also reported. A high dispersion across industries clearly emerges as a stylized fact. The highest OLS sectoral coefficient (health and social work) is almost 3 times higher than the lowest (Other community, social and personal service activities, Private households with employed persons, Extra-territorial organizations and bodies).

[Table 2 here]

Comparing OLS and QR coefficients, it can be noted that the industry with the highest (lowest) value in terms of OLS *is not* the industry with the highest (lowest) spread between the top and the bottom of the distribution, both for the θ90-θ10 and θ75-θ25 differences.

Similarly to M - who compares only the θ90-θ10 quantile difference in returns between the 15 high-paying and 15 low-paying industries (ranked in terms of OLS) - we find that the interquantile differences between the extremes of the distribution are not bigger for the industries with high OLS returns compared to the industries with low OLS returns. With regard to the θ90-θ10 spread, results evidence that the gap between the 6 industries with higher OLS returns and their 6 counterparts with lower estimated OLS coefficients is tighter and *only equal* to 0.007. The same difference was about 0.026 in M. Using the same OLS ranking of industries, such gap is found to be *even negative* for the θ75-θ25 spread. According to M, for the unobserved heterogeneity to be a critical element for wage premia determination, we would expect a really bigger differential between the top of the distribution, where the highest ability workers are expected to be, and its bottom.

5. Robustness check

Further, we test whether gaps between quantile coefficients estimated in our QR are statistically significant. The test has been carried out with respect to the 2 spreads considered in the paper (θ90-θ10=0 and θ75-θ25=0) and to all quantiles. More specifically, p-values are obtained through a bootstrapped variance-covariance matrix that includes between quantile blocks. Similarly to M, the results indicate that in most cases the 2 linear hypothesis (θ95-θ5=0 and θ75-θ25=0) are found to be not significant. As expected, significance decreases when the interquantile spread also decreases. Finally, in addition to M, we find that even the joint equality of coefficients at all quantiles is not rejected in many cases.

Moreover and similarly to M, while the OLS returns are well correlated with almost all the QR returns (table 3), both in terms of Pearson and Spearman's correlations, they are not correlated with the difference between the top and the bottom of the distribution. The Pearson's correlation coefficient between OLS and the θ75-θ25 quantile difference is even negative.

[Table 3 here]

6. Conclusions

In this paper we have applied OLS and QR techniques to the last 2007 EU-SILC wave in order to clear up a possible puzzling presence of inter-industry inequality wage premia for full-time male workers in Portugal. In doing so, we have updated results achieved by Martins (2004) for the year 1995.

We have found high dispersion of OLS returns amongst the 12 considered sectors of the Portuguese economy, but a weak correlation between OLS and the magnitude of 2 quantile benchmarking differences (θ95-θ5 and θ75-θ25). Furthermore, in accordance to Martins (2004) we have demonstrated that if sectors are ranked by OLS coefficients and divided in 2 groups (high and low returns), the interquantile difference is small (and for the difference θ75-θ25 is even negative). Tests on the 2 linear hypothesis θ95-θ5=0 and θ75-θ25=0 are found to be not significant while OLS are well correlated with almost all the QR coefficients but not with the difference between the top and the bottom of the distribution.

These results suggest that for Portuguese adult male workers, unobserved heterogeneity possibly generated by individual skills is an irrelevant matter in the wage determination, while other factors such as the relative strength of industries in a small country may play a more important role.

References

Gibbons, R. and L. Katz, 1992, Does Unmeasured Ability Explain Inter-Industry Wage Differentials?, Review of Economic Studies 59, 515–535.

Hartog, J., Pereira, P.T., and J.C. Vieira, 2000, Inter-industry wage dispersion in Portugal: high but falling. Empirica 27, 353–364.

Koenker, R. and G. Bassett, 1978, Regression quantiles. Econometrica 46, 33– 50.

Krueger, A. and L. Summers, 1988, Efficiency Wages and the Inter-Industry Wage Structure, Econometrica 56, 259–293.

Martins P., 2004, Industry Wage Dispersion: Evidence from the Wage Distribution, Economics Letters, 83(2), 157-163.

Tables

Tab. 1. Summary statistics

Industries - NACE 1.1 Class.	EU-SILC class.	%
Fishing, Agriculture, hunting and forestry	a+b	7.17
Mining and quarrying, Manufacturing, Electricity, gas and water supply	c+d+e	22.96
Construction	f	19.42
Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household good	g	15.83
Hotels and restaurants	h	4.49
Transport, storage and communication	i	5.67
Financial intermediation	j	2.31
Real estate, renting and business activities	k	4.81
Public administration and defence, compulsory social security	l	10.03
Education	m	3.13
Health and social work	n	1.86
Other community, social and personal service activities, Private households with employed persons, Extra-territorial organizations and bodies	o+p+q	2.31
Total		100

Tab. 2. Conditional returns to schooling - OLS and QR - and significance tests

	a+b	c+d+e	f	g	H	i	j	k	l	m	n	o+p+q
OLS	0.047	0.076	0.068	0.047	0.068	0.068	0.061	0.064	0.081	0.098	0.116	0.039
	0.0805	*0.0000*	*0.0000*	*0.0000*	*0.0003*	*0.0000*	*0.0075*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0325*
θ=0.10	0.044	0.032	0.034	0.023	0.021	0.032	0.015	0.027	0.052	0.096	0.081	0.026
	0.0128	*0.0026*	*0.0142*	*0.0731*	*0.2342*	*0.1863*	*0.3397*	*0.1699*	*0.0046*	*0.0000*	*0.1018*	*0.4696*
θ=0.25	0.039	0.051	0.045	0.040	0.033	0.068	0.032	0.052	0.085	0.102	0.104	0.060
	0.0007	*0.0022*	*0.0003*	*0.0229*	*0.2408*	*0.0881*	*0.1080*	*0.0090*	*0.0000*	*0.0000*	*0.0000*	*0.0659*
θ=0.50	0.031	0.079	0.053	0.041	0.086	0.061	0.071	0.072	0.093	0.098	0.112	0.045
	0.1303	*0.0000*	*0.0000*	*0.0000*	*0.0003*	*0.0155*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0313*
θ=0.75	0.065	0.093	0.090	0.047	0.089	0.086	0.106	0.065	0.093	0.089	0.126	0.053
	0.1943	*0.0000*	*0.0000*	*0.0038*	*0.0000*	*0.0002*	*0.0011*	*0.0231*	*0.0000*	*0.0000*	*0.0003*	*0.0031*
θ=0.90	0.073	0.089	0.100	0.068	0.071	0.088	0.055	0.059	0.069	0.047	0.146	-0.002
	0.3177	*0.0000*	*0.0000*	*0.0008*	*0.0003*	*0.0000*	*0.1385*	*0.0917*	*0.0000*	*0.0411*	*0.0000*	*0.9356*
θ90-θ10	0.029	0.057	0.066	0.045	0.050	0.056	0.040	0.032	0.017	-0.049	0.065	-0.028
	0.6136	*0.0000*	*0.0306*	*0.0250*	*0.2051*	*0.0357*	*0.3716*	*0.2950*	*0.1537*	*0.0486*	*0.1612*	*0.6064*
θ75-θ25	0.025	0.042	0.045	0.007	0.057	0.019	0.074	0.013	0.008	-0.013	0.023	-0.007
	0.407	*0.012*	*0.008*	*0.642*	*0.148*	*0.209*	*0.023*	*0.571*	*0.29*	*0.472*	*0.525*	*0.842*
All θ eq.	*0.6816*	*0.0000*	*0.0009*	*0.1464*	*0.0080*	*0.0165*	*0.0000*	*0.4531*	*0.0013*	*0.3864*	*0.3115*	*0.2876*

Note. Data are from cross sectional UDB SILC 2007 – version 1 of March 2009; p-values in italics.

Tab. 3. Correlation between OLS and QR coefficients

	Pearson	Spearman
θ=0.10	0.782	0.664
θ=0.25	0.778	0.624
θ=0.50	0.903	0.885
θ=0.75	0.812	0.711
θ=0.90	0.633	0.448
θ90-θ10	0.074	0.265
θ75-θ25	-0.048	0.004