© *Investigaciones Regionales, 24 – Pages 59 to 90*
Section ARTICLES

Investigaciones Regionales

Overeducation and its effects on wages: a closer look at the Spanish regions

Laura Hernández *, Lorenzo Serrano **

ABSTRACT: This paper uses data from the 2004 to 2009 Living Conditions Survey (LCS) to analyze the wage gap between the adjusted and the overqualified employees in the Spanish regions using standard Mincer equations, quantile regression and the Oaxaca-Blinder decomposition. The results indicate that in Spain there is a 28% difference between the gross hourly wage between the overqualified and well-matched employees, of which 25 percentage points correspond to the discrimination effect and only three percentage points correspond to the characteristics of the individuals and the firms they work in. These results show that the effects of overeducation on the regional economies are genuine and substantial and present a considerable heterogeneity.

JEL Classification: J24, J31, R23.

Keywords: Overeducation, education mismatch, returns to education, quantile regression, regional labour markets.

Sobreeducación y sus efectos sobre los salarios: una mirada a las regiones españolas

RESUMEN: Este trabajo utiliza datos de la Encuesta sobre Condiciones de Vida (ECV) desde 2004 hasta 2009 para analizar la brecha salarial entre los trabajadores ajustados y los sobrecualificados en las regiones españolas utilizando ecuaciones de Mincer estándar, regresiones cuantílicas y la descomposición de Oaxaca-Blinder. Los resultados indican que en España hay una diferencia del 28% entre el salario bruto por hora que reciben los trabajadores sobrecualificados y los adecuadamente ajustados, de los cuales 25 puntos porcentuales se deben al efecto discriminación y únicamente tres puntos al efecto de las características de los individuos y de las empresas donde trabajan. Estos resultados muestran que los efectos de la sobree-

* Ivie.

** Universitat de València&Ivie.

Correspondent autor: Lorenzo Serrano, email: *Lorenzo.Serrano@uv.es*, mailing address: Departamento de Análisis Económico, Facultad de Economía, Universitat de València, Avda. dels Tarongers s/n, 46022-Valencia, Spain.

Lorenzo Serrano wishes to thank the Spanish Ministry of Science and Innovation for its financial support (ECO2011-23248).

Received: April 24th 2012 / Accepted: June 15th 2012.

ducación sobre las economías regionales son genuinos y sustanciales y presentan una considerable heterogeneidad.

Clasificación JEL: J24, J31, R23.

Palabras clave: Sobreeducación, desajuste educativo, rendimientos a la educación, mercados de trabajo regionales.

Introduction

Overeducation is becoming one of the main economic issues in Spain. This is a serious problem from an individual point of view but also from an aggregate one as it seems that an important part of the working population has acquired more education than required in their respective jobs. Overeducation is related to inefficiency problems, as individuals may not take advantage of all the knowledge and skills and education is costly not only for the individuals but also to the whole society. The existence of overeducation in the modern economies shows that it is not enough to improve the skills and competences of the labour force, but it is necessary also to achieve a better match between education and jobs. This mismatch may reduce productivity and competitiveness, both fostering underemployment and unemployment. Thus, the main consequences of overeducation are personal dissatisfaction and lower wages for workers and less productivity for firms. All this may lead to a lower volume of production related to the inputs used, so that these differences with respect to other economies or firms can cause losses in competitiveness. This is especially true at a regional level since regions are fully open economies extremely dependent on their competitiveness. Therefore, the way each regional economy deals with this problem is a key factor to understand their final economic performance. However, the analysis must consider not only the frequency of overeducation but also its specific effects on productivity looking at the wage differentials across regions and their determinants.

In this paper, we use data from the Spanish Living Conditions Survey (LCS) for the period 2004-2009 to analyze the wage differentials caused by overeducation in the Spanish regions. We use a quantile regression approach to account for the existence of non-uniform effects of overeducation over the wage distribution. Thus, we can measure the changes in salary between the overeducated and the adjusted workers at different deciles of the wage distribution and conclude if there is a larger dispersion among overeducated workers. There is also a part of the overeducation literature that relates unobservable characteristics of the individuals, like innate ability or motivation, to different locations in the wage distribution. In that case, more skilled workers would be located at the upper part or the wage distribution and vice versa. With quantile regression we can estimate the wage differentials between low and high ability overeducated workers and see if, related to the believe that overeducated individuals are less able and productive, their lower wages are a consequence or their lower innate abilities and are not related to the mismatch between education and skills required in their jobs. According to this idea workers located in the lower part of the wage distribution and thus less able should show a greater wage differential. However, like Budría and Moro-Egido (2007, 2009) for the EU countries, we find that overeducated employees in the upper part of the wage distribution are more penalized than overeducated workers in the lower part of the wage distribution. This in turn, as these authors point out, seems to be the consequence of a lack of efficiency in the allocation of skills to the characteristics of the productive sector.

We also investigate whether the wage gap between the overeducated and the adjusted workers in the Spanish regions is due to differences in the individuals' and firms' characteristics or to differences in the way the Spanish regional labour markets compensate these characteristics. We do it through the Oaxaca-Blinder decomposition.

The paper is organized as follows. In Section 1 we present some overeducation approaches and evidences and the approximation followed in this paper to measure it. In Section 2 we present the dataset and estimate the average regional effect of overeducation on wages. In Section 3 we present the quantile regression model and its results. In section 4 we present the Oaxaca-Blinder decomposition and its results. Section 5 concludes.

1. Overeducation background

Overeducation is related to the possession of an education level higher than the one required by the job. Although education can be measured differently, the evidence is that there is an important part of overeducated workers. The existence of overeducation contradicts the Human Capital Theory (Becker, 1964), as this theory predicts that workers with higher education levels will be paid higher wages and we find workers with the same level of education and significant wage differentials. This evidence could be rationalized within the HCT framework if educational mismatches were found to be a short run singularity (Sicherman, 1991; Alba-Ramírez, 1993), but there seems to be evidence supporting that workers remain overeducated for much longer time spells (Robst, 1995; Rubb, 2003; Dolton and Vignoles, 2000; McGuinness, 2003).

Concerning international evidence, the differential between the overeducated and the matched workers ranges between 12% (Dolton and Vignoles, 2000), 18% (Dolton and Silles, 2003) or 27% (Chevalier, 2003) for the United Kingdom, 13% for the United States (Verdugo and Verdugo, 1989) or 11% (Cohn and Kahn, 1995), 26% for the Netherlands (Groot, 1993) or 8% in Portugal (Kiker *et al.,* 1997). There are also some authors supporting the argument that overeducated workers may incorporate less innate skills or abilities in their wage gaps. Others (Groot, 1996) find that the wage penalty for overeducated workers is related to tenure, which means that as time goes by the employers find out the real productivity of the workers and discriminate those with fewer abilities than the qualifications they possess.

There are also other theories besides Human Capital Theory that explain the existence of overeducated workers. Carrier Mobility Theory (Galor and Sicherman, 1990) considers an initial situation of overeducation for workers, which will gain experience and specific skills to gain higher occupation levels over time to finally match their qualification level to their occupation. Signaling Theory (Spence, 1973) assumes the existence of an excess of education as a signal to employers in order to be hired. Credential Hypothesis (van der Meer and Wielers, 1996) relies on the difficulty of the employers to measure the true individual's productivity, thus using the educational credentials as a strong proxy for the potential worker's productivity. In that case, individuals may acquire an excess of education in order to fight for a job with other candidates. Job Competition Theory (Thurow, 1975) states that unemployed candidates for a given job are ranked in a hypothetical queue, so overeducation could be an optimal response in order to improve or maintain their position in the queue in order to get the job. Matching Theory (Jovanovic, 1979) assumes the existence of job's misallocations by the existence of search costs and imperfect information.

One of the main difficulties arising from the analysis of overeducation is how to measure it. There are mainly three different methods commonly used. Some of them are based on the systematic evaluation of the jobs and their specific requirements (objective measures), others rely on the subjective perception of the workers about their potential mismatch (subjective measures) and the third type of overeducation measurement is based on empiric analysis (statistic measures). As usual, all of them have advantages and disadvantages but the most frequent decision criterion is the availability of data and information. The objective measures rely on a meticulous analysis of the jobs based on their difficulty, main characteristics, education requirements or special abilities. In this case, the comparison of the education level of the workers and the characteristics of the jobs will determine whether a mismatch exists. This kind of analysis requires an exhaustive analysis. Even more, this process requires a continuous update of the list of occupations given the quick absorption of new technologies and its consequences on the jobs' characteristics. Moreover, not all the countries have such detailed and disaggregated information in order to perform this analysis. The OECD proposes an approximation for an objective and comparable measure of overeducation based on the ISCO classification of occupations and the ISCED classification of education and a correspondence between the occupations and the education level required, but this solution has also some limitations as it relies on the homogeneity of the educational profiles and the occupations between countries. If differences between countries in the requirements of the jobs are important, which is something very likely when working with a high level of aggregation in the definition of the different occupations, and the education systems are different, then the use of a common classification could be misleading.

The subjective method relies on the individuals' opinion about their own mismatch and the overeducated present usually higher dissatisfaction levels in the workplace. It is important to notice that there is a type of overeducation commonly accepted among individuals when they are satisfied with their jobs although they are not well matched with their formation because they usually consider education as a personal consumption good (from a vocational point perspective) or a social status component, but not an investment. There is also a trend towards an upward bias of overeducation, as individuals may exaggerate the requirements of their jobs (Peiró and Montalvo, 2008).

The statistical method is based on the observed adjustment within each occupation through the observation of the most frequent education level. Some authors, like Verdugo and Verdugo (1989), prefer to take the mean of the years of education within each of the occupations, so that individuals above one standard deviation of the average years of education in a given occupation will be classified as overeducated and vice versa. However, these methods are based on the arbitrariness of using the standard deviation as a threshold for detecting overeducation and the assumption that the occupation-education mismatch follows a normal distribution. In that case, if an occupation group had a high incidence of overeducated individuals, then the mean would be affected by this phenomenon thus underestimating its composition (Dolton and Vignoles, 2000). Furthermore, according to Sicherman (1991), the classification of occupations may group jobs for which the educational requirements differ substantially and this effect may be reinforced as we move to a greater occupational aggregation. The statistical method is also very sensitive to the labour market situation (Hartog, 2000), in the sense that if there is a surplus of qualified workers then workers with a higher qualification than the one required for their jobs will be hired, so there would be an underestimation of the overeducation effect and viceversa. The first effect could arise in regions or countries in periods of general education improvements. Alternatively, other authors (Kiker *et al.,* 1997) prefer to use the mode as an alternative to the mean. In that case, individuals with an education level higher than the modal within each occupation group will be overeducated and vice versa. This definition is supposed to be less sensitive to the existence of outliers in the educational distribution.

Unlike in any other Spanish analysis related to overeducation in which subjective or statistical measures are used (Alba, 1993; Beneito *et al.,* 1996; Alba and Blázquez, 2004; Budría and Moro-Egido, 2008; García-Montalvo and Peiró, 2009; Nieto and Ramos, 2010) or the analysis focuses on young workers (Rahona, 2008), in this paper we will use the objective method proposed by the OECD matching the educational and the occupational levels of the workers based on the data provided by the Spanish Living Conditions Survey . In this survey there are no subjective statements related to the education mismatch and regarding the regional analysis we preferred to use a more objective measure of overeducation consistent over time and less dependent on conjunctural circumstances rather than a statistical one like the mean or the mode. Other methods, like the overeducation, required education, and undereducation (ORU) rely on the years of education required to perform specific occupation, so that overeducation is defined as the surplus of years and undereducation implies the opposite. We use a similar approach relying on dummy variables for the different possible matching between education levels and occupations instead of years of education and center the attention on the group of employees which might be overeducated, thus not taking into account undereducation.

The ISCO (International Standard Classification of Occupations) produced by the International Labour Organization can be used to distinguish the different qualifications and skills related to the education levels required to perform the jobs grouped by this classification (tables 1 and 2). The 1-digit educational and occupational groups are classified as high-skilled, intermediate or low-skilled depending on the capacities and abilities related to them and, finally, a correspondence table between occupations and education levels results from matching them together (table 3).

Table 1. Conversion of ISCO-88 9 categories to 3 categories

	Low-skilled	*Intermediate*	*High-skilled*
1. Legislators, senior officials and managers			x
2. Professionals			x
3. Technicians and associate professionals			x
4. Clerks		x	
5. Service and sales workers		x	
6. Skilled agricultural, forestry and fishery workers		x	
7. Craft and related trades workers		x	
8. Plant and machine operators, and assemblers		x	
9. Elementary occupations	x		

Source: OECD (2007).

Table 2. Conversion from ISCED 7 categories to 3 categories

	Low-skilled	*Intermediate*	*Skilled or highly skilled*
0. Pre-primary education or preschool	x		
1. Primary education	x		
2. Lower secondary education	x		
3. Upper secondary education		x	
4. Post-secondary non-tertiary education		x	
5. First stage of tertiary education			x
6. Second stage of tertiary education			x

Source: OECD (2007).

Table 3. Correspondence between ISCED education level and ISCO employment level

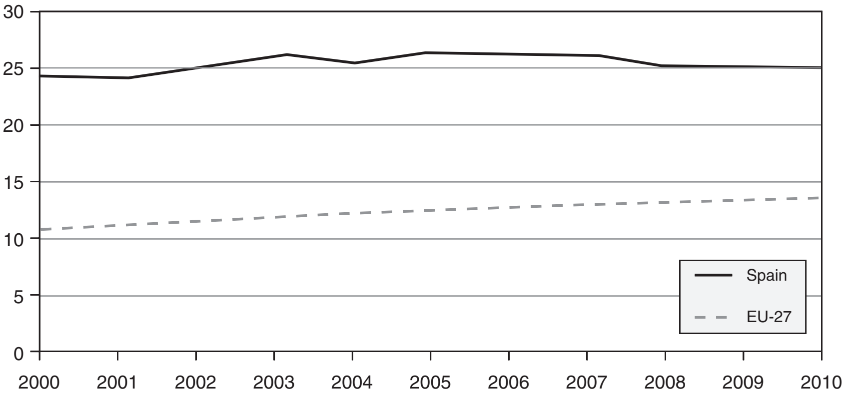
		ISCO employment level		
		Low-skilled	*Intermediate*	*Skilled or highly skilled*
ISCED education level	Low-skilled	Adjusted	Under-qualified	Under-qualified
	Intermediate	Over-qualified	Adjusted	Under-qualified
	Skilled or highly skilled	Over-qualified	Over-qualified	Adjusted

Source: OECD (2007).

From the previous table we can describe three different categories from the adjustment between the occupation groups and the education levels: overqualification, well-matched workers and underqualification. Overqualification arises for high education (ISCED 5-6) and intermediate or low-skilled occupations (ISCO 4-9) or intermediate education levels (ISCED 3-4) and low-skilled occupations (ISCO 9).As the purpose of this paper is to analyze the overeducation phenomenon, we will not take into account the population with low-skilled levels of education (ISCED 0-2) because this group cannot be overqualified. Thus, we will analyze the population with education levels ISCED 3-4 (post-secondary education) and ISCED 5-6 (high education) which will be either adjusted or overqualified.

Following the OECD methodology for the overeducation measurement we can observe that the percentage of overeducated workers with medium and high education is around 25% in Spain, whereas this percentage is much smaller for the EU (around 13%). In that sense, Spain is a special case in the European Union with one of the highest levels of overeducation jointly with Cyprus and followed by Ireland

Figure 1. Overeducation. Spain and EU-27, 2000-2010 (% of medium and high education workers)



Source: Eurostat and own calculations.

and the UK. The Spanish case could be explained in part by the fact that ISCED 5B is rarely found to be matched with ISCO 1-3 occupations. However, if we analyze overeducation in Spain only for employees with a university degree (ISCED 5A-6), this percentage sinks to 20%.

2. Data and methodology

We use data from the 2004-2009 waves of the Spanish Living Conditions Survey (SLCS). The SLCS is a new source of statistical information in the Community environment that strengthens the current European statistical infrastructure in order to respond to the Commission's needs to obtain initial information on the distribution of income and social exclusion in Europe, and acts as a base for the formulation of its social policy in different spheres, and to monitor the effects of these policies in the whole of the European Union (EU). Between 1994 and 2001, the European Union Household Panel (EUHP) survey satisfied these political needs. Nevertheless, given the need to update its content in view of the new demands, and to improve its functioning (especially as regards the speed at which data is produced), the EUHP was replaced with the SLCS. Although persons of all ages are part of the target population, not all persons are eligible to respond to the individual questionnaire. The population under investigation (target population) is persons who are members of private households who live in main family dwellings, as well as said households. Although persons of all ages are part of the target population, not all persons are exhaustively researched since the only persons who can be selected for exhaustive investigation are those aged 16 or over on December 31st of the year prior to the interview. For each region (Autonomous Community) an independent sample that represents it is designed, due to one of the objectives of the survey being to facilitate regional data.

This survey contains personal and labour market characteristics, such as wage, hours worked, age, activity sector, occupation, if the employee supervises other employees, firm size, marital status and nationality, among other variables. Potential labour experience may be estimated as the difference between the age of the employee and the age at which he declared to have finished his studies. Individuals are asked to report the maximum level of education that they have completed according to three categories based on the ISCED-97 classification (OECD, 2003): less than upper secondary (ISCED 0-2), upper secondary (ISCED 3-4) and tertiary education (ISCED 5-6).Our sample consists of employees between 17 and 64 years old not working for the armed forces and reporting a maximum level of education equivalent to ISCED 3-4 or ISCED 5-6 since workers with ISCED 0-2 cannot be overeducated according to our measure of overeducation. The hourly wage is calculated by dividing the gross monthly earnings by the hours usually worked per week multiplied by 4. Since we pool the observations for all the years, a set of dummies controls for a possible time trend and wages are corrected for inflation. Furthermore, we use sample weights provided by the SLCS to make the sample comparable to the population; all the standard errors are robust to heteroskedasticity arising from the sample design. We account for all the Spanish regions except for Ceuta and Melilla.

Table 4. Distribution of the employees of the SLCS survey by educational occupation groups classified by skills. Spain, 2004-2009

		Occupation by skill level			*Total*
		Low-skilled	*Intermediate*	*Skilled or highly skilled*	
ISCED education level	Low-skilled	11.57	26.43	1.68	39.68
	Intermediate	3.16	16.84	3.96	23.95
	Skilled or highly skilled	1.19	13.16	22.02	36.37
	Total	15.91	56.43	27.66	100.00

Source: SLCS and own calculations.

In Table 5 we report some descriptive statistics. Data from the SLCS confirm the existence of significant wage differentials between the Spanish regions, with mean hourly wages ranging from 9.1 euros in Canarias to 13.2 euros in País Vasco, which are almost 50% higher. This regional pattern is coherent with other statistic sources and usual among other economic development indicators. Madrid and the regions located in the north-east occupy the top positions in gross hourly wages, which is also what the SLCS shows relative to the educational levels: the percentage of employees with high education (ISCED 5-6) is higher than 45% in Madrid (45.9%), Navarra (47.3%) and País Vasco (53.4%). On the other hand, Murcia, Canarias and Baleares show percentages of less than 30%. Comunidad Valenciana, Extremadura, Andalucía and Castilla-La Mancha have an amount of high educated employees near 30%. Other variables such as age or labour experience show more homogeneity.

Also the job's characteristics are different between regions, as the percentage of employees in high-skilled occupations (ISCO 1-3) is around 40% in Madrid and 35% in País Vasco, whereas in other regions like Murcia or Canarias this percentage is 22%. Also the percentage of non-skilled jobs (ISCO 9) is very high in Murcia (24%) and doubles the weight of these occupations in other regions. We can also observe that in Madrid, Cataluña and Aragon, jobs which entangle supervision tasks are much frequent than in Murcia, Galicia or Extremadura. Finally, the firm's characteristics constitute a fundamental dimension in which inequalities are apparent between regions. Madrid, País Vasco, Cantabria and Navarra show a higher presence of bigger firms (40%-50%), whilst in Murcia big firms barely arrive to 22%.

For this particular sample of the SLCS for the 2004-2009 period we find that a third of the employees with medium and high education are overeducated. This percentage is almost 37% in Navarra and 21.7% in Baleares. One of the main drawbacks of the education-occupation matching following the methodology proposed in table 3 is the fact that it does not allow to relate specific degrees to specific occupations as it aims to capture this matching in a broader sense in order to enable international comparability. This should not be the case when analyzing a single country, or, as in our case, a regional performance, because of the occupation-education existing homogeneity. In our case, the availability of data did not allow us to go any further, as our database disaggregates occupation in the 9 commonly ISCO main-groups and the

Table 5. Descriptive statistics. Pool 2004-2009

	Gross monthly wage (euros)	*Gross hourly wage (euros)*	*Potential experience (years)*	*Age (years)*	*ISCED 5-6 (%)*	*ISCED 3-4 (%)*	*ISCO 1-3 (%)*	*ISCO 9 (%)*	*1-19 employees (%)*	*50 or more employees (%)*	*Supervision tasks (%)*	*Part-time (%)*	*Overeduc. (%)*	*Obs.*
Andalucía	1,543	10.0	20.1	38.9	33.0	20.9	26.3	20.6	34.5	28.4	19.6	11.5	27.6	6,334
Aragón	1,790	11.5	22.2	41.3	38.2	26.1	27.6	13.4	29.6	37.8	26.0	10.1	32.0	2,830
Asturias	1,754	11.2	21.8	41.2	36.8	28.9	26.5	12.1	34.4	38.4	23.2	8.9	27.0	2,366
Baleares	1,712	11.1	21.3	39.9	29.1	24.2	26.2	14.0	35.4	30.7	24.7	8.9	21.7	2,440
Castilla y León	1,673	10.8	21.5	40.7	40.4	23.8	26.6	15.6	32.5	38.1	20.0	9.0	32.4	3,645
Castilla-La Mancha	1,605	10.5	20.7	39.3	33.0	21.2	25.0	16.3	37.2	27.4	20.8	8.7	29.4	2,877
C. Valenciana	1,536	9.7	21.0	39.1	30.0	24.1	25.1	14.3	35.2	30.3	23.9	10.6	28.1	5,421
Canarias	1,409	9.1	20.7	40.8	27.7	22.8	22.3	21.6	35.2	28.6	20.7	9.0	28.2	3,141
Cantabria	1,728	11.2	21.7	40.2	39.9	26.2	27.8	14.4	27.4	41.7	20.1	11.0	30.4	1,545
Cataluña	1,808	11.6	20.9	39.0	36.2	27.3	30.6	14.9	29.8	37.7	27.5	11.0	26.1	7,510
Extremadura	1,400	9.2	20.6	39.2	31.5	19.9	25.0	23.6	39.8	22.1	18.5	11.0	27.7	2,168
Galicia	1,480	9.4	21.6	40.2	33.9	22.7	23.8	14.6	35.1	31.7	18.3	8.9	30.3	3,975
La Rioja	1,544	10.1	21.7	40.2	35.7	23.5	22.5	15.2	33.8	30.8	18.8	9.4	34.4	2,072
Madrid	1,964	12.3	20.5	41.0	45.9	26.0	39.7	13.3	23.8	50.3	30.6	8.7	24.5	5,056
Murcia	1,501	9.5	20.7	38.8	26.6	25.0	21.6	24.3	33.9	30.2	18.0	9.4	30.1	2,777
Navarra	1,914	12.6	20.1	39.5	47.3	22.8	27.2	10.3	28.7	41.9	24.1	13.4	36.8	2,537
P.Vasco	1,993	13.2	21.6	41.6	53.4	20.5	34.8	12.1	25.0	44.3	24.7	10.9	34.2	3,430
Total	**1,675**	**10.8**	**21.0**	**40.0**	**36.4**	**23.9**	**27.7**	**15.9**	**32.1**	**35.0**	**23.0**	**10.1**	**29.0**	**60,124**
Dispersion range	**593**	**4.1**	**2.1**	**2.8**	**26.7**	**8.9**	**18.1**	**13.9**	**16.0**	**28.2**	**12.5**	**4.8**	**15.0**	
Standard Deviation	**186**	**1.2**	**0.6**	**0.9**	**7.3**	**2.5**	**4.5**	**4.1**	**4.4**	**7.3**	**3.6**	**1.3**	**3.8**	

Source: SLCS and own calculations.

education levels are also grouped in three main categories according to the 0-2, 3-4 and 5-6 ISCED levels. Table 4 shows the distribution of the three educational levels of the SLCS for the years 2004 to 2009 by skilled groups.

As a preliminary exercise we estimate the effect of the worker's and the workplace's characteristics on the gross hourly wage, adding also a dummy variable for the effect of being overeducated, using OLS techniques through a Mincerian (Mincer, 1974) semi-logarithmic wage equation:

$$y = \alpha + X'\beta + over\ \gamma + \varepsilon \qquad (1)$$

where α is the intercept term, X is a vector of independent variables measuring a range of individual and job characteristics such as marital status, level of education (equal to one for high education and equal to zero for post-secondary education), gender, potential labour market experience calculated as the difference between the year of the survey and the year that the individuals finished their formal studies, a second degree polynomial in labour market experience, supervising other employees, part-time work, etc, ε is the residual term and there is also dummy variable (*over*) that takes the value of one if the worker is overeducated and zero if the worker is well-matched. Since we pool the observations for all the years, we use a set of dummies controls for a possible time trend. The dependent variable is the logarithm of the average hourly wage of each employee. We also apply the Gardeazabal and Ugidos (2004) identification restriction in order to obtain the estimation effects for the omitted reference for all the categorical variables. The results for the subset of employees with upper secondary or tertiary education, including also occupational categories in column 1, are shown in table 6. Ceteris paribus, gross hourly wages increase with the educational level, the employee's potential experience and the firm's size and is also positively associated with performing supervision tasks and working in skilled occupations (ISCO 1-3). On the contrary it is significantly smaller for women, foreign workers and part-time workers. Finally, the economic activity and the region have a significant influence on wages (with higher wages in the financial sector, education or health, and regions such as País Vasco and Navarra).

Column 2 presents the results for the estimation including the dummy variable for overeducation, defined following the correspondence between the education level and the occupation group (table 3), thus excluding the ISCO dummies for the occupation groups. The coefficients are very similar for all variables to the ones estimated in column 1, although it is important to remark the positive increase of education, reflecting the fact that the occupation level depends positively on the education of the individuals, which is also an important component of the returns to education. The effect of being overeducated is significant and negative, reflecting ceteris paribus a decrease of –25% on the gross hourly wage, a very high cost associated with this problem. If mismatched employees had jobs suited to their qualifications there would be a substantial increase in their productivity and given the high levels of overqualification in Spain, this would also increase the national and regional labour productivity. The time dummies included in the regression and all the econometric

Table 6. OLS regression for the log hourly wage

	Total (1)		*Total (2)*	
ISCED 5-6	0.101	***	0.318	***
Experience	0.015	***	0.015	***
Experiencie2	0.000	***	0.000	***
Female	–0.111	***	–0.102	***
Foreign	–0.117	***	–0.128	***
Married	0.090	***	0.095	***
Supervision	0.123	***	0.154	***
Part–time work	–0.048	**	–0.052	**
1-10 employees	–0.082		–0.089	
11-19 employees	–0.035	***	–0.028	**
19-49 employees	0.031	***	0.027	**
50 or more employees	0.086	***	0.089	***
Agriculture, cattle & fishing	–0.121		–0.130	
Extr. ind., manuf, prod. & distrib.	–0.026	*	–0.046	***
Construction	–0.055	***	–0.073	***
Trade & repair	–0.071	***	–0.103	***
Hotels & restaurants	–0.076	***	–0.110	***
Transportation, storage & repair	–0.043	**	–0.046	**
Financial intermediation	0.183	***	0.191	***
Real estate, renting & business act.	–0.074	***	–0.072	***
Public adm. & defence; comp. ss.	0.160	***	0.166	***
Education	0.149	***	0.229	***
Health & vet. act, social service	0.068	***	0.083	***
Other social activities, etc	–0.096	***	–0.090	***
ISCO 1: Managers	0.351			
ISCO 2: Professionals	0.286	***		
ISCO 3: Technicians & ass. professionals	0.066	***		
ISCO 4: Clerical support workers	–0.046	***		
ISCO 5: Service and sales workers	–0.115	***		
ISCO 6: Skilled agric., forestry and fishery workers	–0.133	**		
ISCO 7: Craft and related trades workers	–0.125	***		
ISCO 8: Plant and machine operators, & assemblers	–0.092	***		
ISCO 9: Elementary occupations	–0.192	***		
Andalucía	–0.069		–0.066	
Aragón	0.032		0.026	
Asturias	–0.002		–0.007	
Baleares	0.104	***	0.117	***
Castilla y León	–0.012		–0.027	
Castilla-La Mancha	0.020		0.024	
C. Valenciana	–0.043	**	–0.033	*
Canarias	–0.079	***	–0.077	***
Cantabria	–0.006		–0.020	
Cataluña	0.045	***	0.049	***
Extremadura	–0.080	***	–0.076	**
Galicia	–0.091	***	–0.097	***
La Rioja	–0.001		–0.003	
Madrid	0.009		0.016	
Murcia	–0.040	**	–0.040	*
Navarra	0.120	***	0.128	***
País Vasco	0.094	***	0.086	***

Table 6. *(cont.)*

	Total (1)		*Total (2)*	
2004	–0.115		–0.115	
2005	–0.068	***	–0.067	***
2006	0.027	***	0.028	***
2007	0.034	***	0.033	***
2008	0.046	***	0.044	***
2009	0.076	***	0.078	***
Overeducated			–0.250	***
Constant	2.041	***	1.997	***
Obs.	34,169		34,169	
R^2	0.541		0.511	
Adjusted R2	0.540		0.511	

*** p<0.01, ** p<0.05, * p<0.1.

analysis performed in the next sections should capture any inflation or trend effects or any break due to the economic crisis after 2007. As it can be seen, as time goes by their values are more positive.

These results are broadly consistent with the empirical literature on this topic for the Spanish case. The wage penalty estimated for self-assessed overeducated workers was –9% according to Budría and Moro-Egido (2007), –17% according to Alba (1993) or, using a strict definition of subjective overeducation, –4% according to Aguilar and García-Crespo (2008).Evidence about a wage penalty for overeducated workers with respect to workers having a job matching their qualifications was also found in Nieto and Ramos (2010). Rahona *et al.* (2010) find that overeducated workers have a relative impact with respect to adjusted workers of –3.2 percentage points on the rate of return of years of schooling for the years of overeducation. In order to compare these results, we must take into account that they are related to different overeducation definitions (self-assessment, mean, mode...), different surveys (PHOGUE, ECBC, EES) or different analytical tools (ORU, panel, quantile...) to the ones used in our paper.

In Spain there are no formal barriers to labour mobility, but it is useful to analyze each regional labour market performance separately. The results for estimating a separate regression for each region as in table 6-column 2 are shown in table 7. Some of them are qualitatively similar: positive effect of education, firm's size and supervision tasks; negative effect on being a woman, a foreigner, etc. However, there are also differences in the magnitude and relevance of the effects between regions. Regional labour markets seem to pay some employee's characteristics such as education, gender, nationality or experience differently. Something similar happens with the job's and firm's characteristics.

In particular, being overeducated has a significant and negative effect over the gross hourly wage in all regions, but this effect is heterogeneous across the Spanish territory. The overeducation penalty is -17.5% in Cataluña, whereas in Castilla-La

Table 7. OLS regressions for the log hourly wage, by region

	Andalucía		*Aragón*		*Asturias*		*Baleares*		*Castilla y León*		*Castilla-La Mancha*		*C. Valenciana*		*Canarias*		*Cantabria*	
ISCED 5–6	0.317	***	0.329	***	0.407	***	0.206	***	0.345	***	0.361	***	0.374	***	0.321	***	0.310	***
Experience	0.014	**	0.019	**	0.022	*	0.012		0.028	***	0.015	*	0.016	***	0.004		0.024	**
Experiencie^2	0.000		0.000		0.000		0.000		0.000	***	0.000		0.000		0.000		0.000	
Female	–0.059	*	–0.177	***	–0.073		–0.041		–0.189	***	–0.092	*	–0.121	***	0.032		0.042	
Foreign	–0.168	*	–0.208	***	–0.355	*	–0.160	*	–0.063		–0.251	***	–0.053		–0.055		0.019	
Married	0.062		0.019		0.124	*	–0.034		0.043		0.049		0.028		0.097	*	0.108	
Supervision	0.133	***	0.079	*	0.190	***	0.239	***	0.183	***	0.129	***	0.076	*	0.208	**	0.299	***
Part-time work	–0.077		–0.065		0.058		0.091		–0.050		0.016		0.038		–0.141	*	–0.125	
1-10 employees	–0.097		–0.071		–0.130		–0.104		–0.082		–0.087		–0.092		–0.173		–0.060	
11-19 employees	–0.034		–0.003		0.001		–0.013		–0.030		–0.075		–0.044		–0.051		0.020	
19-49 employees	0.042		–0.040		0.058		0.085	*	0.023		0.039		–0.009		0.105	**	–0.070	
50 or more employees	0.090	***	0.115	***	0.071	*	0.032		0.089	***	0.123	***	0.145	***	0.118	***	0.109	**
Agriculture, cattle & fishing	–0.154		0.167		–0.112		–0.016		0.050		0.012		–0.425		–0.015		–0.381	
Extr. ind., manuf, prod. & distrib.	–0.096	*	–0.072		0.082		0.096		0.002		–0.092		–0.088	*	0.005		–0.058	
Construction	–0.075		–0.057		–0.014		–0.231	**	0.004		–0.058		–0.041		–0.155	**	–0.074	
Trade & repair	–0.135	***	–0.101	**	–0.115	*	–0.162	*	–0.054		–0.023		–0.081		–0.179	***	0.006	
Hotels & restaurants	–0.136	*	–0.175	***	–0.297	***	–0.214	***	–0.267	***	–0.146	*	0.094		–0.076		–0.250	***
Transportation, storage & repair	–0.104		–0.050		–0.125	*	0.057		–0.163	***	–0.163	**	0.007		–0.278	***	–0.171	**
Financial intermediation	0.363	***	0.054		0.139		0.114		0.325	***	0.134		0.329	***	0.354	***	0.571	**
Real estate, renting & business act.	0.010		0.026		–0.165	**	0.036		–0.164	***	–0.091		–0.094	**	–0.067		–0.129	
Public adm. & defence; comp. ss.	0.174	***	0.140	*	0.178	**	0.102		0.188	***	0.234	***	0.202	***	0.121	**	0.232	***
Education	0.229	***	0.147	**	0.183	**	0.209	***	0.318	***	0.314	***	0.230	***	0.283	***	0.385	***
Health & vet. act, social service	0.028		0.084		0.071		0.225	***	0.084		0.193	***	0.097		0.201	***	–0.016	
Other social activities, etc.	–0.106	*	–0.164	**	0.177		–0.215	**	–0.324	**	–0.315	**	–0.230	***	–0.195	*	–0.115	
2004	–0.114		–0.112		–0.088		–0.089		–0.155		–0.102		–0.119		–0.139		–0.181	
2005	–0.064	***	–0.027		–0.060	*	–0.103	***	–0.059	***	–0.042		–0.078	***	–0.054	**	–0.089	**
2006	0.026		0.015		0.028		0.015		0.009		0.023		–0.001		0.034		0.154	***
2007	0.032	**	0.002		0.005		0.040		0.055	**	0.005		0.039	**	0.025		0.037	
2008	0.051	***	0.033		0.013		0.032		0.052	***	0.006		0.060	***	0.052	**	0.023	
2009	0.070	***	0.090	***	0.101	***	0.105	***	0.098	***	0.111	***	0.099	***	0.083	***	0.057	*
Overeducated	–0.236	***	–0.286	***	–0.325	***	–0.182	**	–0.249	***	–0.343	***	–0.269	***	–0.267	***	–0.244	***
Constant	1.948	***	2.115	***	1.838	***	2.181	***	1.899	***	1.947	***	1.961	***	1.950	***	1.782	***
N	3,197		1,745		1,471		1,238		2,200		1,501		2,783		1,528		987	
R^2	0.568		0.550		0.567		0.568		0.636		0.679		0.552		0.680		0.588	
Adjusted R^2	0.564		0.542		0.559		0.558		0.631		0.673		0.547		0.674		0.576	

*** p<0.01, ** p<0.05, * p<0.1.

Table 7. (cont.)

	Cataluña		*Extremadura*		*Galicia*		*La Rioja*		*Madrid*		*Murcia*		*Navarra*		*País Vasco*	
ISCED 5–6	0.237	***	0.405	***	0.340	***	0.293	***	0.369	***	0.285	***	0.261	***	0.272	***
Experience	0.011	**	0.015		0.020	***	0.002		0.014	**	0.021	**	0.025	***	0.015	**
Experiencie^2	0.000		0.000		0.000		0.000		0.000		0.000		0.000	*	0.000	
Female	–0.092	***	–0.142	**	–0.157	***	–0.193	***	–0.088	***	–0.164	***	–0.075	*	–0.169	***
Foreign	–0.185	***	–0.279		0.166	*	–0.077		–0.146	**	–0.334	***	–0.086		0.125	
Married	0.110	***	0.150	**	0.113	**	0.033		0.133	***	0.228	***	0.005		0.117	***
Supervision	0.175	***	0.141		0.224	***	0.067		0.188	***	0.094	**	0.052		0.106	**
Part-time work	–0.140	***	–0.052		0.023		–0.047		0.038		–0.064		–0.061		–0.005	
1-10 employees	–0.078		–0.124		–0.034		–0.056		–0.119		–0.046		–0.078		–0.107	
11-19 employees	–0.007		0.025		–0.051		–0.075		–0.049		–0.037		–0.170	***	0.014	
19-49 employees	0.025		0.038		–0.029		0.010		0.086	**	0.019		0.091	**	0.034	
50 or more employes	0.060	***	0.061		0.015	***	0.121	***	0.081	***	0.064	**	0.156	***	0.059	**
Agriculture, cattle & fishing	–0.300		0.394		–0.188		–0.193		–0.117		0.084		–0.163		–0.230	
Extr. ind., manuf, prod. & distrib.	–0.029		–0.149	**	–0.072		–0.180	***	0.020		–0.090	*	–0.045		–0.031	
Construction	–0.061		–0.243	**	–0.051		–0.018		–0.109	**	–0.175	***	0.032		–0.038	
Trade & repair	–0.078	*	–0.004		–0.023		–0.196	***	–0.091		–0.170	**	–0.109	**	–0.091	
Hotels & restaurants	–0.090		–0.013		–0.242	***	–0.026		–0.102		–0.027		–0.154	*	–0.028	
Transportation, storage & repair	–0.014		–0.094		–0.011		–0.086		0.015		–0.143	**	0.167	**	–0.119	*
Financial intermediation	0.212	***	0.512	***	0.107		0.322	***	0.102	**	0.317	***	0.102		0.210	***
Real estate, renting & business act.	–0.042		–0.654	***	–0.054		0.112		–0.089	**	–0.105	**	–0.159	*	–0.066	
Public adm. & defence; comp. ss.	0.187	***	0.172	**	0.295	***	0.106	**	0.123	***	0.126	**	0.252	***	0.125	**
Education	0.216	***	0.234	**	0.329	***	0.293	***	0.133	***	0.232	***	0.191	*	0.250	***
Health & vet. act, social service	0.039		0.252	***	0.051		0.051		0.109	**	–0.029		–0.002		0.209	***
Other social activities, etc.	–0.041		–0.408	***	–0.140	*	–0.186	*	0.007		–0.021		–0.111		–0.191	**
2004	–0.078		–0.106		–0.165		–0.122		–0.134		–0.151		–0.113		–0.078	
2005	–0.064	***	–0.078	**	–0.096	***	–0.076	***	–0.069	***	–0.094	***	–0.065	***	–0.042	**
2006	0.010		0.018		0.051	**	0.012		0.062	***	0.055	**	–0.028		0.026	
2006	0.028	**	0.050	*	0.058	**	0.039		0.036	**	0.020		0.086	***	0.001	
2007	0.044	***	0.012		0.059	***	0.047	*	0.044	***	0.063	**	0.030		0.031	*
2008	0.060	***	0.105	***	0.093	***	0.099	***	0.061	***	0.107	***	0.091	***	0.062	***
2009	–0.078		–0.106		–0.165		–0.122		–0.134		–0.151		–0.113		–0.078	
Overeducated	–0.175	***	–0.287	***	–0.303	***	–0.195	***	–0.253	***	–0.297	***	–0.217	***	–0.263	***
Constant	2.112	***	1.872	***	1.790	***	2.107	***	1.950	***	1.963	***	2.117	***	2.145	***
N	4,477		1,090		2,106		1,134		3,218		1,381		1,708		2,405	
R^2	0.433		0.632		0.548		0.670		0.464		0.704		0.521		0.467	
Adjusted R^2	0.429		0.622		0.541		0.661		0.460		0.698		0.513		0.460	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Mancha this wage gap is -34.3%. In some cases, the effect of being overeducated almost compensates (País Vasco, Galicia, Castilla-La Mancha, Baleares, Aragón) or even exceeds (Murcia) the effect of attaining higher levels of education over upper secondary education. These results show that in order to assess the regional overeducation differences in Spain, it is not enough to consider the amount of mismatched workers or their weight on the total employment. Its ultimate impact in terms of productivity and economic performance may be influenced by the characteristics of the workers and the regional productive structure. The same percentage of overeducation may be more or less important depending on these factors. The following sections will analyze these issues.

3. Quantile regression approach

The quantile regression method allows to estimate the impact of overeducation not only on the mean of the wage distribution as in the standard OLS technique, but to estimate the possible different impacts of overeducation over different points of the wage distribution, thus helping us to understand the existence of differences in returns among employees with different levels of unobserved ability, as all of them hold an equivalent observable ability (education level). For the technique to be useful we require there to be sufficient variation in the levels of the exogenous variables across the quantiles in order to obtain statistically significant estimations. The empirical results suggest that the data used here is sufficient to meet that condition.

The QR model can be formally written as follows (see Buchinsky, 1994):

$$\ln\, w_i = X_i\beta_q + \varepsilon_{qi} \;\; with\; Quant_q(\ln\, w_i|X_i) = X_i\beta_q \tag{2}$$

Where X_i is the vector of exogenous variables and β_q is the vector of parameters. $Quant_q(\ln\, w_i|X_i)$ denotes the *qth* conditional quantile of *ln* given *X*. The *qth* regression quantile, $0 < q < 1$, is defined as a solution to the problem:

$$\min_{\beta\in R^k}\left\{\Sigma\rho_q(\ln\, w_i - X_i\beta_q)\right\} \tag{3}$$

where the check function $\rho_q(z) = q_z$ if $z \geq 0$ or $\rho_q(z) = (q-1)z$ if $z < 0$. This problem is solved using linear programming methods, where standard errors for the vector of coefficients are obtained using the bootstrap method described in Buchinsky (1998). It must be noted that if the underlying model were a location model, that is, changes in the explanatory variables producing changes only in the location, not in the shape, of the conditional wage distribution, then all the slope coefficients would be the same for all *p*.

This analysis will allow us to determine if it is unobservable characteristics such as innate ability of personal motivation which are affecting earnings and productivity or it is more due to situations of apparent educational mismatch. If unobservable characteristics were important, the negative effect of being overeducated should be bigger in the lower part of the wage distribution where other kind of abilities could not compensate this mismatch. If we observe the opposite phenomenon, then we could conclude that unobservable characteristics are not as relevant as the pure education-occupation mismatch.

Tables 8 and 9 show the gross hourly wages for the well-matched and the overeducated employees across the wage distribution in Spain and the Spanish regions for the workers with at least upper secondary education (ISCED 3-6) and the workers with high education (ISCED 5-6). For both groups the wage gap grows as we move to higher percentiles. As it can be seen, the range of variation of wages between regions is higher as we move to the upper part of the wage distribution. These differences are slightly smaller for workers with tertiary education for higher deciles but, anyway, they are also higher in the lower part of the wage distribution.

Table 8. Wage differences across the wage distribution. Spain. Pool 2004-2009 (euros 2009)

Deciles	*Overqualif. (ISCED 3-6)*	*Adjusted (ISCED 3-6)*	*Overqualif. / Adjusted (%)*	*Overqualif. (ISCED 5-6)*	*Adjusted (ISCED 5-6)*	*Overqualif. / Adjusted (%)*
10	5.33	6.14	86.78	5.54	8.08	68.49
20	6.22	7.49	83.09	6.51	10.11	64.43
30	7.01	8.84	79.38	7.36	11.96	61.55
40	7.79	10.19	76.48	8.22	13.80	59.54
50	8.66	11.73	73.86	9.13	15.56	58.72
60	9.65	13.62	70.85	10.18	17.45	58.33
70	10.89	15.83	68.78	11.50	19.45	59.14
80	12.58	18.70	67.29	13.28	21.97	60.42
90	15.67	22.83	68.63	16.41	26.01	63.08
Diff. 90-10	10.33	16.68	61.94	10.87	17.93	60.64

As a next step we analyze whether the matched-mismatched wage differential remains stable across the wage distribution controlling for the individual's and firm's characteristics. Table 10 presents the coefficients for the overeducation dummy from the 10^{th} to the 90^{th} percentiles of hourly income. The results for Spain (last row of table 10) confirm the negative effect across the wage distribution, which is more intensive in the highest percentiles (this effects is –19.2% in the 10^{th} percentile and –30.1% in the 90^{th} percentile, a difference of 11 percentage points). This pattern is similar in the Spanish regions except for Cantabria, and the 90^{th}-10^{th} difference is higher than 18 percentage points in Galicia, La Rioja and País Vasco.

Table 11 shows the overeducation coefficients obtained for the employees with tertiary education.The results are similar than the ones shown in table 10 but the

Table 9. Gross hourly wage for the overeducated and matched employees across the wage distribution

A) ISCED 3-6										
	10	20	30	40	50	60	70	80	90	90-10
Overeducated										
Andalucía	5.0	5.8	6.5	7.1	7.9	8.7	9.9	11.6	14.3	9.3
Aragón	6.0	7.0	7.8	8.4	9.3	10.2	11.7	13.2	16.8	10.8
Asturias	4.9	6.2	6.8	7.6	8.9	10.0	11.4	13.2	16.8	11.9
Baleares	5.7	6.7	7.3	7.9	8.8	9.6	11.4	12.6	15.0	9.3
C. y León	5.3	6.1	7.0	7.8	8.6	9.8	11.1	13.0	15.7	10.4
C.-La Mancha	5.2	6.2	7.1	7.8	8.7	9.5	10.9	12.5	15.0	9.8
C. Valenciana	5.2	6.0	6.6	7.2	7.8	8.6	9.5	10.9	13.1	7.9
Canarias	4.7	5.5	6.1	6.7	7.2	8.0	8.7	9.8	12.1	7.5
Cantabria	5.0	6.1	7.0	7.6	8.5	9.6	10.9	13.0	17.5	12.4
Cataluña	5.6	6.8	7.4	8.4	9.5	10.7	11.9	13.7	17.1	11.5
Extremadura	4.7	5.3	6.1	6.7	7.4	8.0	9.1	10.4	14.1	9.4
Galicia	4.6	5.4	6.2	6.8	7.4	8.2	9.1	10.5	12.7	8.1
La Rioja	5.4	6.2	7.0	7.6	8.2	9.1	9.7	11.0	13.2	7.8
Madrid	5.5	6.4	7.2	8.2	9.0	10.0	11.5	13.4	16.9	11.4
Murcia	5.2	6.0	6.6	7.1	7.9	8.7	9.6	10.7	13.5	8.3
Navarra	6.3	7.5	8.4	9.3	10.4	11.3	12.8	14.7	16.8	10.4
País Vasco	6.4	7.8	8.6	9.5	10.4	11.4	12.8	14.5	17.8	11.5
Range. Var.	1.7	2.5	2.5	2.8	3.2	3.4	4.1	4.9	5.7	5.0

B) ISCED 5-6										
	10	20	30	40	50	60	70	80	90	90-10
Overeducated										
Andalucía	5.3	6.1	6.9	7.6	8.4	9.3	10.5	12.5	15.5	10.2
Aragón	6.4	7.3	8.1	8.9	9.8	10.8	12.0	13.9	17.5	11.1
Asturias	5.3	6.4	7.0	8.0	9.3	10.3	12.0	14.0	17.9	12.6
Baleares	6.0	6.9	7.6	8.3	9.3	10.4	11.9	13.2	16.3	10.4
C. y León	5.5	6.3	7.3	8.2	9.0	10.4	11.7	13.6	16.0	10.5
C.-La Mancha	5.5	6.5	7.3	8.2	9.0	10.0	11.3	13.1	15.3	9.8
C. Valenciana	5.4	6.2	7.0	7.5	8.3	9.1	10.1	11.6	13.9	8.4
Canarias	4.9	5.6	6.2	6.8	7.5	8.3	9.2	10.3	12.7	7.8
Cantabria	5.5	6.3	7.3	8.1	8.9	10.0	11.7	14.3	18.0	12.5
Cataluña	5.9	6.9	7.9	9.0	10.2	11.4	12.6	14.6	18.2	12.3
Extremadura	4.8	5.6	6.3	6.9	7.4	8.1	9.5	11.2	15.0	10.2
Galicia	4.7	5.6	6.3	6.9	7.6	8.5	9.5	10.9	13.1	8.4
La Rioja	5.6	6.5	7.2	7.9	8.6	9.4	10.2	11.4	14.1	8.5
Madrid	5.8	6.9	7.9	8.8	9.7	10.9	12.3	14.4	18.0	12.2
Murcia	5.4	6.2	6.8	7.5	8.7	9.4	10.2	11.3	14.3	8.9
Navarra	6.5	7.8	8.7	9.7	10.6	11.7	13.1	15.1	17.0	10.5
País Vasco	6.8	8.0	8.8	9.7	10.6	11.8	13.1	14.9	18.2	11.3
Range. Var.	2.1	2.4	2.6	3.0	3.2	3.7	4.0	4.8	5.5	4.8

A) ISCED 3-6										
	10	20	30	40	50	60	70	80	90	90-10
Matched										
Andalucía	5.9	7.1	8.5	9.9	11.5	13.3	15.6	18.6	22.4	16.5
Aragón	6.7	8.0	9.3	10.7	12.3	14.3	16.3	18.6	22.7	16.0
Asturias	5.8	7.0	8.4	9.8	11.4	13.4	15.9	19.0	23.5	17.7
Baleares	6.7	8.0	9.3	10.8	12.4	14.7	16.9	19.5	23.2	16.5
C. y León	6.2	7.5	8.7	10.3	11.7	13.5	15.8	18.7	22.5	16.3
C.-La Mancha	6.1	7.5	8.8	10.0	11.6	13.7	16.0	19.3	23.8	17.6
C. Valenciana	5.8	6.9	8.1	9.3	10.5	12.3	14.5	17.1	20.3	14.5
Canarias	5.3	6.3	7.3	8.6	10.0	12.0	14.8	17.6	22.0	16.7
Cantabria	6.1	7.4	8.7	10.6	12.4	14.4	16.2	19.0	23.9	17.8
Cataluña	6.8	8.3	9.3	10.6	12.1	13.9	16.0	18.9	23.0	16.2
Extremadura	5.3	6.8	8.2	9.6	11.3	13.3	15.6	19.0	22.9	17.6
Galicia	5.3	6.4	7.5	8.8	10.2	12.0	14.2	17.0	21.8	16.5
La Rioja	6.2	7.0	8.2	9.4	10.4	11.9	13.9	16.6	21.5	15.3
Madrid	6.5	8.1	9.7	11.2	12.9	14.6	16.7	19.7	24.3	17.9
Murcia	5.6	7.0	8.2	9.2	10.6	12.4	14.5	17.3	21.1	15.4
Navarra	7.2	8.6	10.0	11.3	12.9	15.0	17.4	20.2	24.1	16.9
País Vasco	7.7	9.2	10.6	12.1	13.9	15.7	17.8	20.9	25.9	18.2
Range. Var.	2.4	2.9	3.3	3.6	3.9	3.7	3.9	4.2	5.7	3.8

B) ISCED 5-6										
	10	20	30	40	50	60	70	80	90	90-10
Matched										
Andalucía	7.8	9.7	11.6	13.6	15.3	17.4	19.3	21.6	24.8	17.0
Aragón	8.9	10.9	12.6	14.6	16.2	17.7	19.3	21.8	25.4	16.5
Asturias	7.7	10.3	12.3	14.4	16.0	17.9	20.2	22.7	27.3	19.6
Baleares	9.8	11.9	13.6	15.7	17.1	18.8	20.8	23.0	27.6	17.8
C. y León	7.7	9.8	11.7	13.5	15.6	17.7	19.9	21.9	25.4	17.7
C.-La Mancha	7.7	9.7	11.4	13.7	15.7	17.6	19.8	22.9	27.1	19.3
C. Valenciana	7.5	9.5	11.1	12.8	14.5	16.2	18.0	19.8	22.6	15.1
Canarias	7.3	9.5	11.1	13.5	15.5	17.3	19.4	21.8	24.9	17.6
Cantabria	7.0	9.8	12.4	14.3	15.9	17.7	19.5	21.9	26.9	19.8
Cataluña	8.6	10.4	11.9	13.8	15.3	17.4	19.4	22.0	26.2	17.6
Extremadura	7.7	9.8	11.8	13.8	15.4	17.6	19.8	22.1	25.8	18.1
Galicia	6.9	8.9	10.7	12.9	14.7	16.3	18.3	21.2	25.4	18.6
La Rioja	7.8	9.8	11.0	12.7	14.4	15.8	18.1	21.1	24.0	16.2
Madrid	8.5	10.5	12.2	13.8	15.6	17.4	19.5	22.5	27.8	19.3
Murcia	8.1	10.0	12.0	13.7	15.4	16.9	18.6	21.0	23.7	15.6
Navarra	9.0	11.1	12.7	14.9	17.3	19.1	20.9	23.6	27.6	18.6
País Vasco	8.9	10.8	12.8	14.4	16.0	18.0	20.6	23.9	27.8	18.9
Range. Var.	2.9	3.0	2.8	3.0	2.9	3.4	2.9	4.1	5.3	4.8

Table 9. *(cont.)*

A) ISCED 3-6											**B) ISCED 5-6**										
	10	20	30	40	50	60	70	80	90	90-10		10	20	30	40	50	60	70	80	90	90-10
Difference between overeducated and matched employees											**Difference between overeducated and matched employees**										
Andalucía	–0.9	–1.4	–2.0	–2.7	–3.6	–4.5	–5.7	–7.0	–8.1	–7.2	Andalucía	–2.5	–3.6	–4.8	–6.0	–6.9	–8.0	–8.8	–9.1	–9.3	–6.9
Aragón	–0.7	–1.0	–1.6	–2.2	–3.1	–4.0	–4.6	–5.3	–5.8	–5.2	Aragón	–2.4	–3.6	–4.5	–5.6	–6.3	–6.9	–7.3	–7.9	–7.8	–5.4
Asturias	–0.9	–0.9	–1.6	–2.2	–2.5	–3.4	–4.5	–5.8	–6.6	–5.8	Asturias	–2.4	–3.9	–5.4	–6.4	–6.7	–7.6	–8.2	–8.7	–9.3	–7.0
Baleares	–0.9	–1.3	–2.1	–2.8	–3.7	–5.0	–5.5	–6.9	–8.2	–7.2	Baleares	–3.8	–5.0	–5.9	–7.4	–7.8	–8.5	–8.9	–9.8	–11.3	–7.5
C. y León	–0.9	–1.3	–1.7	–2.5	–3.2	–3.6	–4.7	–5.7	–6.8	–5.9	C. y León	–2.2	–3.5	–4.4	–5.3	–6.5	–7.3	–8.3	–8.3	–9.4	–7.2
C.-La Mancha	–0.9	–1.2	–1.7	–2.2	–2.9	–4.2	–5.1	–6.8	–8.8	–7.9	C.-La Mancha	–2.2	–3.2	–4.1	–5.6	–6.7	–7.6	–8.5	–9.8	–11.7	–9.5
C. Valenciana	–0.6	–1.0	–1.5	–2.1	–2.7	–3.8	–5.0	–6.2	–7.1	–6.5	C. Valenciana	–2.1	–3.3	–4.1	–5.2	–6.2	–7.2	–7.9	–8.3	–8.7	–6.6
Canarias	–0.6	–0.7	–1.2	–1.9	–2.8	–4.0	–6.0	–7.8	–9.9	–9.2	Canarias	–2.4	–3.9	–4.9	–6.7	–8.0	–9.0	–10.2	–11.5	–12.2	–9.7
Cantabria	–1.0	–1.3	–1.8	–3.1	–4.0	–4.7	–5.3	–6.0	–6.4	–5.4	Cantabria	–1.5	–3.5	–5.1	–6.3	–6.9	–7.7	–7.8	–7.6	–8.9	–7.3
Cataluña	–1.2	–1.5	–1.9	–2.2	–2.6	–3.2	–4.1	–5.2	–5.9	–4.7	Cataluña	–2.7	–3.4	–4.0	–4.8	–5.1	–6.0	–6.8	–7.5	–8.0	–5.3
Extremadura	–0.6	–1.5	–2.1	–2.9	–3.9	–5.2	–6.6	–8.6	–8.8	–8.2	Extremadura	–2.9	–4.2	–5.5	–6.9	–7.9	–9.5	–10.3	–10.9	–10.8	–7.9
Galicia	–0.7	–0.9	–1.3	–2.0	–2.8	–3.8	–5.1	–6.5	–9.1	–8.4	Galicia	–2.1	–3.3	–4.4	–5.9	–7.1	–7.8	–8.8	–10.3	–12.3	–10.2
La Rioja	–0.7	–0.8	–1.2	–1.8	–2.2	–2.8	–4.2	–5.6	–8.3	–7.6	La Rioja	–2.2	–3.3	–3.8	–4.8	–5.8	–6.4	–7.9	–9.7	–9.9	–7.7
Madrid	–1.0	–1.7	–2.5	–3.0	–3.8	–4.5	–5.2	–6.3	–7.4	–6.5	Madrid	–2.7	–3.6	–4.3	–5.0	–5.8	–6.5	–7.2	–8.1	–9.8	–7.1
Murcia	–0.4	–1.0	–1.6	–2.1	–2.7	–3.6	–4.9	–6.6	–7.5	–7.1	Murcia	–2.7	–3.8	–5.1	–6.2	–6.7	–7.5	–8.4	–9.7	–9.4	–6.6
Navarra	–0.9	–1.0	–1.6	–2.0	–2.5	–3.6	–4.6	–5.5	–7.3	–6.4	Navarra	–2.5	–3.3	–4.0	–5.1	–6.7	–7.5	–7.8	–8.5	–10.6	–8.1
País Vasco	–1.3	–1.3	–2.0	–2.6	–3.4	–4.3	–4.9	–6.4	–8.1	–6.7	País Vasco	–2.1	–2.8	–4.0	–4.8	–5.4	–6.2	–7.5	–9.0	–9.6	–7.6
Range. Var.	0.9	1.0	1.3	1.2	1.7	2.4	2.5	3.4	4.0	4.6	Range. Var.	2.3	2.2	2.1	2.7	2.9	3.5	3.5	4.0	4.5	4.9

magnitude of the effect is much higher now. For Spain, the effect of being overeducated over the gross hourly wage is –20.9% in the 10th percentile and –36.4% in the 90th percentile, which amounts to a difference of 15.5 percentage points. In this case Cantabria is also a special case and does not follow the general pattern of intensification of the wage penalty across the wage distribution. Differences between the 90th and the 10th percentile are higher than 20 percentage points in Galicia, Andalucía and País Vasco. Moreover, the overeducation effect for the 90th percentile is higher than 40% in five regions and no smaller than 26% for all of them.

These results suggest that the importance of the unobservable characteristics is not the main issue concerning the wage differentials related to overeducation: the overeducation indicator is reflecting a genuine mismatch with significant effects on the regional productivity.

4. Oaxaca-Blinder decomposition

The Oaxaca-Blinder decomposition (Blinder 1973; Oaxaca, 1973) divides the wage differential between two groups of workers (by sex, race, etc.) in a counterfactual framework. This decomposition allows disentangling the part of the differential due to differences in the characteristics of the individuals (which is usually captured by simple OLS regression) and the part due to differences in the way the labour mar-

Table 10. Quantile regression coefficients for the dummy «overeducated». Employees with ISCED 3-4 or ISCED 5-6 education

	q10	*q20*	*q30*	*q40*	*q50*	*q60*	*q70*	*q80*	*q90*	*q90-q10*
Andalucía	–0.202	–0.228	–0.242	–0.247	–0.260	–0.280	–0.301	–0.314	–0.341	–0.139
Aragón	–0.222	–0.241	–0.223	–0.225	–0.232	–0.232	–0.235	–0.247	–0.229	–0.007
Asturias	–0.248	–0.229	–0.257	–0.270	–0.278	–0.269	–0.320	–0.360	–0.397	–0.149
Baleares	–0.185	–0.263	–0.249	–0.260	–0.308	–0.303	–0.297	–0.324	–0.306	–0.121
Castilla y León	–0.202	–0.226	–0.218	–0.261	–0.270	–0.281	–0.285	–0.288	–0.302	–0.100
Castilla-La Mancha	–0.222	–0.192	–0.241	–0.293	–0.299	–0.315	–0.354	–0.336	–0.295	–0.073
C. Valenciana	–0.200	–0.205	–0.216	–0.207	–0.233	–0.229	–0.229	–0.238	–0.216	–0.016
Canarias	–0.129	–0.199	–0.228	–0.242	–0.265	–0.261	–0.293	–0.278	–0.303	–0.173
Cantabria	–0.253	–0.247	–0.309	–0.312	–0.334	–0.299	–0.303	–0.256	–0.248	0.005
Cataluña	–0.207	–0.204	–0.224	–0.223	–0.214	–0.233	–0.241	–0.236	–0.228	–0.020
Extremadura	–0.270	–0.298	–0.304	–0.349	–0.359	–0.350	–0.340	–0.376	–0.374	–0.104
Galicia	–0.205	–0.236	–0.270	–0.275	–0.314	–0.320	–0.338	–0.336	–0.395	–0.190
La Rioja	–0.127	–0.192	–0.216	–0.208	–0.226	–0.225	–0.218	–0.246	–0.309	–0.182
Madrid	–0.191	–0.214	–0.226	–0.268	–0.277	–0.284	–0.293	–0.298	–0.331	–0.141
Murcia	–0.132	–0.193	–0.191	–0.227	–0.232	–0.241	–0.236	–0.247	–0.275	–0.143
Navarra	–0.239	–0.227	–0.230	–0.239	–0.232	–0.228	–0.255	–0.278	–0.272	–0.033
País Vasco	–0.098	–0.169	–0.200	–0.212	–0.238	–0.234	–0.252	–0.275	–0.282	–0.183
Spain	–0.192	–0.214	–0.230	–0.245	–0.260	–0.268	–0.280	–0.293	–0.301	–0.110

Note: All coefficients are significant at the 0.01 level.

Table 11. Quantile regression coefficients for the dummy «overeducated». Employees with ISCED 5-6 education

	q10	*q20*	*q30*	*q40*	*q50*	*q60*	*q70*	*q80*	*q90*	*q90-q10*
Andalucía	–0.217	–0.215	–0.270	–0.282	–0.299	–0.319	–0.354	–0.390	–0.440	–0.223
Aragón	–0.255	–0.251	–0.263	–0.242	–0.259	–0.265	–0.267	–0.279	–0.263	–0.008
Asturias	–0.243	–0.186	–0.285	–0.299	–0.309	–0.301	–0.316	–0.363	–0.415	–0.172
Baleares	–0.244	–0.333	–0.322	–0.334	–0.325	–0.322	–0.349	–0.394	–0.350	–0.105
Castilla y León	–0.241	–0.247	–0.246	–0.280	–0.292	–0.320	–0.346	–0.367	–0.364	–0.123
Castilla-La Mancha	–0.279	–0.219	–0.281	–0.336	–0.332	–0.391	–0.426	–0.439	–0.451	–0.172
C. Valenciana	–0.226	–0.261	–0.258	–0.225	–0.252	–0.284	–0.303	–0.311	–0.308	–0.082
Canarias	–0.182	–0.203	–0.260	–0.243	–0.261	–0.264	–0.306	–0.338	–0.374	–0.192
Cantabria	–0.266	–0.344	–0.358	–0.340	–0.337	–0.314	–0.333	–0.283	–0.218*	0.048
Cataluña	–0.186	–0.219	–0.238	–0.240	–0.240	–0.255	–0.289	–0.289	–0.283	–0.097
Extremadura	–0.334	–0.363	–0.389	–0.438	–0.477	–0.466	–0.498	–0.471	–0.445	–0.110
Galicia	–0.257	–0.292	–0.295	–0.318	–0.357	–0.380	–0.417	–0.426	–0.488	–0.232
La Rioja	–0.160	–0.254	–0.223	–0.235	–0.245	–0.256	–0.259	–0.264	–0.356	–0.197
Madrid	–0.208	–0.209	–0.226	–0.284	–0.283	–0.291	–0.299	–0.310	–0.265	–0.057
Murcia	–0.212	–0.274	–0.327	–0.321	–0.336	–0.321	–0.320	–0.298	–0.305	–0.092
Navarra	–0.255	–0.241	–0.218	–0.250	–0.240	–0.242	–0.280	–0.285	–0.307	–0.051
País Vasco	–0.112	–0.156	–0.201	–0.212	–0.234	–0.243	–0.258	–0.266	–0.319	–0.206
Spain	–0.209	–0.236	–0.252	–0.271	–0.290	–0.304	–0.320	–0.339	–0.364	–0.155

Note: Coefficients are significant at the 0.01 level unless specified otherwise. ** $p<0.05$, * $p<0.1$

ket values these characteristics. This last part could be accounted as a measure for discrimination or differences due to unobserved predictors.

This decomposition needs two groups (we will use «O» for overeducated and «M» for well-matched employees), an outcome variable Y (the log hourly wage), and a set of predictors relating individual and firm characteristics (with the same specification as for the initial OLS analysis from the previous section: education, sex, immigrant condition, activity sector, etc.). This decomposition tries to answer the question on how much of the mean outcome difference $R = E(Y_M) - E(Y_O)$, where $E(Y)$ denotes the expected value of the log hourly wage, is accounted for by group differences in the predictors based on the linear model:

$$Y_\mu = X_\mu \beta_\mu + \varepsilon_\mu,\ E(\varepsilon_\mu) = 0,\ \mu \in \{A,\ B\} \tag{4}$$

where X is a vector containing the predictors and a constant, β contains the slope parameters and the intercept, and ε is the error:

$$R = E(Y_M) - E(Y_O) = E(X_M)'\beta_M - E(X_O)'\beta_O \tag{5}$$

since $E(\beta_\mu)$ and $E(\varepsilon_\mu) = 0$, which can be rearranged as follows (see Jann, 2008):

$$R = [E(X_M) - E(X_O)]'\beta_O + E(X_O)'(\beta_M - \beta_O) + [E(X_M) - E(X_O)]'(\beta_M - \beta_O) \tag{6}$$

This is a «three-fold» decomposition, as the differential can be divided in three components: $R = E + C + I$.

The first component amounts to the part due to group differences in the predictors (the «endowments effect»). The second component measures the contribution of differences in the coefficients (including differences in the intercept). The third component is an interaction term which takes into account the existence of simultaneous differences in both endowments and coefficients.

The previous decomposition is formulated from the viewpoint of the overeducated group (O). We can see that the group differences in the predictors are weighted by the coefficients of the overeducated to determine the endowments effect (E). In other words, the E component measures the expected change in the overqualified's mean outcome, if overeducated had the same characteristics of the well-matched (M) employees. Moreover, for the second component (C) the differences in coefficients are weighted by the overeducated mean characteristics. That is, the second component measures the expected change in the overeducated's mean outcome, if they had the well-matched coefficients (betas). Thus, the differential can also be expressed from the viewpoint of the well-matched (reverse three-fold decomposition):

$$R = [E(X_M) - E(X_O)]'\beta_M + E(X_M)'(\beta_M - \beta_O) - [E(X_M) - E(X_O)]'(\beta_M - \beta_O) \tag{7}$$

In order to account for both possibilities we take the average and the interaction effect disappears:

$$R = \tfrac{1}{2}R + \tfrac{1}{2}R = \tfrac{1}{2}\Big[[E(X_M) - E(X_O)]'\ \beta_O + E(X_O)'(\beta_M - \beta_O) + [E(X_M) - E(X_O)]'\ (\beta_M - \beta_O)\Big] + \tfrac{1}{2}\Big[[E(X_M) - E(X_O)]'\ \beta_M + [E(X_M)'\ (\beta_M - \beta_O) - [E(X_M) - E(X_O)]'\ (\beta_M - \beta_O)\Big] \quad (8)$$

In the previous sections we have found that the situations of apparent mismatch between the education levels of the employees and the requirements of their occupations are very common in the Spanish regions. We have also found that there is a substantial impact of overeducation in the labour productivity of the Spanish regions and this overeducation seems to be a situation of genuine mismatch and not the result of unobservable characteristics of the employees.

The Oaxaca-Blinder decomposition allows us to further analyze the factors determining the negative effect of overeducation in the Spanish regions. Results from OLS regressions (tables 5 and 6) showed the importance of several personal characteristics, as well as characteristics of the types of job and the firms, and the different influence of these variables across the Spanish regions. With this technique we can separate out the wage differential between the overeducated and the well-matched employees into differences in the average individual's and job's characteristics between these two groups of workers and differences in the returns to these characteristics for each region.

Table 12 shows the results of the decomposition for workers with at least upper secondary education. As it can be seen, differences between the overeducated and the well-matched workers are mainly due to differences of the effect of these characteristics in the wage determination for these two types of workers (as much as 90% is due to the coefficients effect), usually through the effect of how experience is valued by the labour market. In Spain, of the 28% difference between overqualified and well-matched employees, 25 percentage points correspond to the coefficients effect, whereas only 3 percentage points correspond to the endowments effect. Only in Baleares, Andalucía, Canarias, Cantabria and La Rioja the endowments effect has some relevance, contributing between a 20% and a 54% to the negative wage effect of being overeducated.

If we analyze only the workers with tertiary education (table 13), the results are similar but not as accentuated. First we can observe that the wage gap is much bigger (44% for Spain instead of 28%). Moreover, the coefficients effect is also more relevant (accounting for a 73% in Spain) than the characteristics effect (27%). We can observe this pattern in all the Spanish regions except for Baleares. However, if we compare these results with the ones in table 12 we can see that the endowments effect is generally bigger and significant and represents more than a quarter of the total wage differential. This is something that can be explained by the fact that matched workers with tertiary education have an ad-

Table 12. Oaxaca decomposition (mean of the Oaxaca and the reverse Oaxaca decomposition). Population with ISCED 3-4 or ISCED 5-6 education. Pool 2004-2009 [1]

	Adjusted (lnw)	*Overqualif. (lnw)*	*Difference*	*Endowments effect*	*Coefficients effect*	*% Endow.*	*% Coeff.*
Andalucía	2.50	2.17	0.33	0.10 **	0.23 ***	30.58	69.42
Aragón	2.55	2.31	0.24	–0.03	0.26 ***	–11.12	111.12
Asturias	2.54	2.32	0.22	–0.07	0.29 ***	–31.58	131.58
Baleares	2.61	2.38	0.23	0.12	0.11 **	54.01	45.99
Castilla y León	2.50	2.28	0.23	–0.01	0.23 ***	–3.43	103.43
Castilla-La Mancha	2.67	2.25	0.42	0.01	0.41 ***	3.02	96.98
C. Valenciana	2.51	2.20	0.31	0.02	0.29 ***	5.35	94.65
Canarias	2.49	2.12	0.37	0.09	0.27 ***	24.90	75.09
Cantabria	2.53	2.23	0.30	0.09	0.20 ***	31.00	69.00
Cataluña	2.56	2.35	0.21	0.03	0.18 ***	16.01	84.00
Extremadura	2.44	2.26	0.17	–0.12	0.30 ***	–70.30	170.29
Galicia	2.39	2.15	0.24	–0.05	0.30 ***	–21.23	121.23
La Rioja	2.49	2.21	0.28	0.06	0.22 ***	22.94	77.06
Madrid	2.61	2.33	0.29	0.03	0.26 ***	10.46	89.54
Murcia	2.51	2.23	0.28	–0.05	0.33 ***	–17.31	117.31
Navarra	2.68	2.46	0.22	–0.01	0.23 ***	–5.09	105.09
País Vasco	2.73	2.43	0.30	0.04	0.26 ***	12.52	87.48
Total	2.56	2.28	0.28	0.03 **	0.25 ***	9.76	90.24

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

vantage with respect to overeducated workers in the sense that they have a higher average potential experience (17 years versus 14.6 years for the overeducated), have higher positions of supervision (39% versus 24% for the overeducated) and part-time work is also less frequent among well-matched workers. We can also observe from tables 12 and 13 that both the characteristics and the coefficients effects tend to generate different results regarding the final effect of overeducation across regions in Spain, thus contributing to regional differences in terms of wages, productivity and economic performance. Generally, the variables that affect the endowments part of the decomposition are potential experience and responsibility and experience has also a very high impact in the coefficients part of the decomposition. These results on the wage penalty for overeducation across regions are in line with previous analyses of the total wage differences across regions across Spanish regions. García and Molina (2002) found a similar effect

[1] Results are similar to a pooled two-fold decomposition between explained and unexplained characteristics.

Table 13. Oaxaca decomposition (mean of the Oaxaca and the reverse Oaxaca decomposition). Population with ISCED 5-6 education. Pool 2004-2009

	Adjusted (lnw)	*Overqualif. (lnw)*	*Difference*	*Endowments effect*	*Coefficients effect*	*% Endow.*	*% Coeff.*
Andalucía	2.71	2.25	0.46	0.20 ***	0.26 ***	44.05	55.95
Aragón	2.81	2.34	0.47	0.09 **	0.38 ***	18.90	81.10
Asturias	2.82	2.37	0.45	–0.07 *	0.52 ***	–15.51	115.51
Baleares	2.85	2.41	0.44	0.30 ***	0.14	67.53	32.47
Castilla y León	2.78	2.29	0.49	0.13 ***	0.36 **	25.83	74.17
Castilla-La Mancha	2.90	2.31	0.59	0.14 ***	0.45 ***	24.14	75.86
C. Valenciana	2.73	2.25	0.49	0.19 ***	0.30 ***	38.45	61.55
Canarias	2.80	2.13	0.67	0.23 ***	0.43 ***	35.12	64.88
Cantabria	2.74	2.26	0.47	0.07	0.41 ***	13.81	86.19
Cataluña	2.76	2.37	0.38	0.10 ***	0.28 ***	25.84	74.16
Extremadura	2.70	2.32	0.38	0.11	0.27 ***	28.12	71.88
Galicia	2.65	2.21	0.44	0.09 **	0.35 ***	20.17	79.83
La Rioja	2.68	2.27	0.40	0.14 ***	0.26 ***	34.44	65.56
Madrid	2.76	2.38	0.38	0.10 ***	0.28 ***	25.55	74.45
Murcia	2.72	2.37	0.34	–0.01	0.36 ***	–4.08	104.08
Navarra	2.80	2.47	0.34	0.01 ***	0.33 ***	2.01	97.99
País Vasco	2.82	2.44	0.38	0.09 ***	0.28 ***	25.15	74.85
Total	2.76	2.33	0.44	0.12 ***	0.32 ***	26.98	73.02

*** p<0.01, ** p<0.05, * p<0.1.

of the endowments and coefficients component. Their results show higher differences for the coefficients effect (remuneration to characteristics) in the North and the East, related to variables with greater influence such as seniority, university level of education, activity sector, supervision tasks and occupation. Motellón *et al.* (2011) find significant regional differences in characteristics and coefficients and their impact on total wage differentials. Following Serrano (2002), these authors suggest that wage differentials due to differences in the returns to characteristics may explain the existence of inefficient regional labour markets. They also argue that a regional homogeneity in the wage distribution could only be possible through a simultaneous equalization of the worker's, firm's and workplaces' characteristics and the returns to these characteristics. A further improvement for further research would be to incorporate a quantile analysis in the Oaxaca-Blinder framework to improve comparisons with the mean and incorporate the whole distribution to compare between regions that show important signs of heterogeneity, especially because of the lack of explicit information in our data source about unobservable ability.

5. Conclusions

This paper has analyzed the overeducation problem and its effects across the Spanish regions based on the SLCS micro data for the period 2004-2009. Our starting point is the evidence drawn from the overeducation results about the Spanish situation both internationally, with a highest level of overeducated workers, and regionally, with evident disparity between regions. We have analyzed the regional overeducation disparities in several stages. First, by means of the estimation of Mincerian wage equations for each region including an overeducation dummy related to skills required for each occupation group. We have found a substantial significant negative effect of overeducation on the gross hourly wage for Spain (–25%) and for each region, ranging from –17.5% to 35%. These results may be associated with serious productivity and low wage problems. Moreover, the fact that this phenomenon has an unequal impact across the Spanish territory has also clear implications regarding regional development.

Secondly, the estimation of quantile regressions confirms the existence of a wage gap across the wage distribution in all regions, thus confirming the relevance of overeducation. Moreover, these findings suggest that the overeducation indicator is capturing genuine mismatch situations beyond the possible effect of unobservable characteristics such as innate ability or motivation, as the wage gaps related to overeducation are systematically more pronounced in the higher percentiles of the wage distribution in practically all regions. This finding rejects the assumption that individuals with lower innate ability, thus more likely located in the lower part of the wage distribution, should be more penalized than the individuals located in the upper part of the wage distribution which could compensate their apparent overeducation with other types of skills.

Finally, an Oaxaca-Blinder decomposition was estimated for all regions in order to distinguish the wage differentials between the overeducated and the matched employees. Results show, as in the previous exercises, that the effect of overeducation is greater for more educated workers (around –44% for tertiary education versus –28% for workers with at least upper secondary education). This wage gap differs from one region to another and seems to be explained by a greater extent by the coefficients effect rather than the endowments (characteristics) effect. Hence, the wage penalty is driven by the way that the labour market values the overeducated or matched worker's characteristics. This is especially evident when we analyze the workers with at least upper secondary education. In this case, characteristics account only for 10% of the wage gap. Their relevance is higher in the case of workers with tertiary education, with contributions close to 25%-30%. Nevertheless, the main part of the wage gap is still due to the different way the labour market values each characteristic.

All these results show that the effects of overeducation on the regional economies are genuine and substantial and present a considerable heterogeneity. The different analyses performed show a common feature: the existence of significant regional differences in the effects of the wage determinants in general and the overeducation wage gap in particular. Hence, overeducation is an additional aspect to be considered when analyzing regional economies and this may imply a first step to analyze it deeper in future research, by controlling the decomposition not only at the mean but for the whole distribution. Even more, circumstances like differences in the regional incidence of temporary work or the relationship between a possible excess of high educated employees and overeducation across regions may constitute plausible explanations of the overeducation issue that should be studied in future analyses. Ultimately, our results show that the regional level is a promising research area in order to better understand the problem of overeducation and the circumstances that may aggravate or mitigate their impact on the economy.

References

Aguilar, M. I., and García-Crespo, D. (2008): «Desajuste educativo y salarios en España: nueva evidencia con datos de panel», *Estadística española,* 50 (168): 393-426.

Alba, A., and Blázquez, M. (2004): «Types of job match, overeducation and labour mobility in Spain», in Büchel, F.; Grip, A., and Mertens, A. (eds.), *Overeducation in Europe: current issues in theory and policy,* Edward Elgar, Cheltenham, UK.

Alba-Ramírez, A. (1993): «Mismatch in the Spanish Labour Market. Overeducation?», *Journal of Human Resources,* 28, 259-278.

Bauer, T. (2002): «Educational mismatch and wages: a panel analysis», *Economics of Education Review*, 21, 221-229.

Becker, G. (1964): *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education,* New York, Columbia UniversityPress.

Beneito, P.; Ferri, J.; Moltó, M. L., and Uriel, E. (1996): «Desajuste educativo y formación laboral especializada: efectos sobre los rendimientos salariales», Documento de trabajo WP-EC 9611, Instituto Valenciano de Investigaciones Económicas.

Blinder, A. S. (1973): «Wage Discrimination: Reduced Form and Structural Estimates», *The Journal of Human Resources,* 8: 436-455.

Buchinsky, M. (1998): «Recent advances in quantile regression models: a practical guideline for empirical research», *Journal of Human Resources,* 33, 88-126.

Budría, S., and Moro-Egidio, A. I. (2007): «Overeducation and Wages in Europe: Evidence from Quantile Regression», *Economic Working Papers,* E2007/04, Centro de Estudios Andaluces, Sevilla.

— (2009): «The Overeducation Phenomenon in Europe», Revista Internacional de Sociología, 67(2), pp. 329-345.

Chevalier, A. (2003): «Measuring mismatch», *Economica,* 70, 509-531.

Cohn, E., and Khan, P. (1995): «The wage effects of overschooling revisited», *Labour Economics,* 2, 67-76.

Dolton, P., and Silles, M. (2001): *Mismatch in the Graduate Labour Market: Some Evidence from Alumni Data,* Centre for the Economics of Education, Discussion Paper Series, 9.

Dolton, P., and Vignoles, A. (2000): «The incidence and effects of mismatch in the UK graduate labour market», *Economics of Education Review,* 19, 179-198.

Duncan, J., and Hoffman, S. (1981): «The incidence and wage effects of overeducation», *Economics of Education Review*, 1(1), 75-86.

Galor, O., and Sicherman, N. (1990): «A theory of Career Mobility», *Journal of Political Economy*, 98(1): 169-192.

García, I., and Molina, J. A. (2002): «Inter-regional wage differentials in Spain», *Applied Economics Letters,* 9: 4, 209-215.

García-Montalvo, J. (1995): «Empleo y sobrecualificación: el caso español». Documento de-trabajo 95-20, FEDEA.

García-Montalvo, J., and Peiró, J. M. (2009): *Análisis de la sobrecualificación y la flexibilidad laboral,* Fundación Bancaja.

Gardeazábal, J., and Ugidos, A. (2004): «More on Identification in Detailed Wage Decompositions», *The Review of Economics and Statistics,* 86(4), 1034-1036.

Green, F.; McIntosh, S., and Vignoles, A. (1999): *Overeducation and Skills: Clarifying the Concepts,* London, Centre for Economic Performance.

Groot, W. (1993): «Overeducation and the returns to enterprise related schooling», *Economics of Education Review,* 17(4), 299-309.

— (1996): «The incidence of, and returns to overeducation in the UK», *Applied Economics,* 28, 1345-1350.

Groot, W., and Van den Brink, H. (2000): «Overeducation in the Labor Market: A Meta-Analysis», *Economics of Education Review,* 19, 149-158.

Hartog, J. (2000): «Mismatch and earnings: where are we, where should we go», *Economics of Education Review,* 19, 131-147.

Hartog, J., and Oosterbeek, H. (1988): «Education, allocation and earnings in the Netherlands: overschooling?», *Economics of Education Review,* 7, 185-194

Jann, B. (2008): «The Blinder-Oaxaca decomposition for linear regression models», *The Stata Journal,* 8 (4): 453-479.

Jovanovic, B. (1979): «Job Matching and the Theory of Turnover», *Journal of Political Economy,* 87, 972-990.

Kiker, B.; Santos, M., and Mendes de Oliveira, M. (1997): «Overeducation and undereducation: evidence for Portugal», *Economics of Education Review,* 16(2), 111-125.

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

McGuiness, S. (2003): «Graduate overeducation as a sheepskin effect: evidence from Northern Ireland», *Applied Economics,* 35, 597-608.

Mincer, J. (1974): *Schooling, Experience and Earnings,* Columbia University Press, New York.

Motellón, E.; López-Bazo, E., and El-Attar, M. (2011): «Regional heterogeneity in wage distributions: evidence from Spain», *Journal of Regional Science,* 51(3): 558-584.

Nieto, S., and Ramos, R. (2010), «Sobreeducación, educación no formal y salarios: Evidencia para España», Documentos de Trabajo FUNCAS, núm. 577, 2010.

Oaxaca, R. (1973): «Male-Female Wage Differentials in Urban Labor Markets», *International Economic Review,* 14: 693-709.

Oaxaca, R., and Ransom, M. (1994): «On discrimination and the decomposition of wage differentials», *Journal of Econometrics,* 61: 5-21.

OECD (2007): *International Migration Outlook 2007,* OECD Publishing.doi: 10.1787/migr_outlook-2007-en.

Pastor, J. M.; Raymond, J. L.; Roig, J. L., and Serrano, L. (2009): *El rendimiento del capital humano en España,* Fundación Bancaja.

Rahona, M. (2008): «Un análisis del desajuste educativo en el primer empleo de los jóvenes», *Revista Principios-Estudios de Economía Política,* 11, 45-67.

Rahona, M.; Murillo I. P., and Salinas, M. M. (2010): «Incidencia del desajuste educativo en el rendimiento privado de la educación en España», ch. 13, pp. 267-284, in Mancebón-Torrubia, M.ª J.; Ximénez de-Embún, D. P.; Gómez-Sancho, J. M.ª, and Giménez Esteban, G. (eds.), *Investigaciones de Economía de la Educación*, 5, vol. 5, Asociación de Economía de la Educación.

Robst, J. (1995): «College Quality and Overeducation», *Economics of Education Review,* 14(3), 221-228.

Rubb, S. (2003): «Overeducation: A Short or Long Run Phenomenon for Individuals?», *Economics of Education Review,* 22, 389-394.
Serrano, L. (2002): «Salarios regionales y dotaciones de capital humano», *Revista de Economía Aplicada,* 28: 23-38.
Sicherman, N. (1991): «Mismatch in the Labor Market», *Journal of Labor Economics,* 9(2), 101-122.
Spence, M. (1973): «Job market signaling», *Quarterly Journal of Economics,* 87(3), 355-374.
Thurow, L. (1975): *Generating Inequality,* New York, Basic Books.
Van der Meer, P., and Wielers, R. (1996): «Educational Credentials and Trust in the Labor Market», *Kyklos,* 49 (1), 29-46.
Verdugo, R., and Verdugo, N. T. (1989): «The impact of surplus schooling on earnings: Some additional findings», *Journal of Human Resources,* 24, 629-695.

Comment on «Overeducation and its Effects on Wages: A Closer Look at the Spanish regions», by Laura Hernández and Lorenzo Serrano

Raquel Llorente Heras *

The aim of this paper is to show that the wage penalty imposed by over-education may be associated with high regional heterogeneity in Spain. Through the development of several models of probabilities, *logits*, quantile regressions and Oaxaca decompositions, the authors find that it is essential to assume a regional perspective in analysing over-education.

This paper is of significant relevance and interest. First, over-education, a topic that is traditionally studied by labour economists, has recently undergone an important update. In Spain, the pioneering papers of Alba-Ramirez (1993) and García-Montalvo (1995) have established that one of the most adverse effects of labour matching is over-education, generally defined as the attainment of a higher level of education or training than the average level required at the workplace. Presently, the studies of Rahona (2008) and Rahona *et al.* (2010) confirm the maintenance of a stable and important educational mismatch in Spain. Alba and Blázquez (2004) expand the existing knowledge about over-education by analysing labour mobility, and Iglesias *et al.* (2010) show that over-education can be linked to the segregation of women, even in workplaces where it is low, such as in the ICT sector.

Second, the new regional economy emphasises the relevance of the development of regional analyses in most modern studies. For example, the study mentions some papers related to the national economy where the regional perspective was basic (García and Molina, 2002; Motellón *et al.*, 2011; and Serrano, 2002).

Although a similar national education system extends to all Spanish regions, educational matching differs from one region to another. This is caused by variations in regional productive structures and unequal regional requirements for the labour force's educational level. Recently, heterogeneity in regional over-education has been based on the existence of different regional production systems and specialisation patterns. However, it has also been revealed that labour mobility is low among Spanish regions, and therefore, the adjustment between labour requirements at the workplace and the educational attainment of workers does not occur through the movement of the labour force or firms.

All of this means that the adjustment of wages, with a penalty for over-educated workers, is very heterogeneous at the regional level. Therefore, we should ask ourselves which failures in the labour market cause the problems of educational mismatching, the lack of worker and firm mobility, production structures that are not

* Universidad Autónoma de Madrid.

capable of suitably contracting the most qualified employees, inefficient wage allocation in regional markets and an excessive temporality that favours labour rotation and/or educational segregation, among others.

It should be emphasised that the wage penalty for over-educated workers estimated by this paper is approximately 25%. However, it is more worrying that, at the regional level, the wage penalty varies from 17.5% to 35%, setting a distance of 17.5 points, precisely twice the minimum. This establishes a certainly unequal labour scenario for Spanish workers with more education depending on their region.

Moreover, the wage penalty is uneven depending on the position of the worker, with the overall distribution of wages being clearly superior in the higher wage quantiles. However, the wage gap associated with over-education is relevant in all regions. In light of these findings, the authors suggest the existence of several unobservable characteristics that can be captured by over-education at the regional level.

Although over-education is an important variable to take into account in the explanation of regional heterogeneity, this paper, which uses Oaxaca decomposition, shows that the wage gap remains strongly unexplained at the regional level. This result reinforces the idea that future studies of the wage penalty should be developed with more accuracy. In many cases, an important part of the explanation for the wage gap is associated with unobservable variables or characteristics that lead economists to assume a certain self-effacement and methodological reflection. Not only must advances in the study of wage differentials at the regional level continue, but the tools for measuring the wage gap must also be improved.

Finally, I would like to end this comment with a personal reflection. One of the effects of the current economic crisis is the increase in unemployment among low-skilled workers and the occupation of low-skill jobs by people with higher educational attainment (Garrido, 2012). From another perspective, workers with higher levels of education have accepted lower-skill jobs in order to become or remain employed (*i. e.,* downward occupational flexibility). Given this pattern, over-education, a clearly countercyclical labour issue, has increased more than expected during the development of the current crisis. For this reason, some questions should be considered. What type of growth will we achieve if the most qualified workers are employed in lower-skill jobs far below their potential? How will they develop their real productivity within this labour allocation?

References

Alba, A., and Blázquez, M. (2004): «Types of Job Match, Overeducation and Labor Mobility in Spain», in Büchel, F.; de Grip, A., and Mertens, A. (eds.), *Overeducation in Europe: Current Issues in Theory, and Policy,* Cheltenham, UK, Edward Elgar.

Alba-Ramírez, A. (1993): «Mismatch in the Spanish Labor Market: Overeducation?», *The Journal of Human Resources,* 28, 259-278.

García, I., and Molina, J. A. (2002): «Inter-regional wage differentials in Spain», *Applied Economics Letters*, vol. 9(4), 209-215.

García Montalvo, J. (1995): *Empleo y sobrecualificación: El caso español,* Documento de Trabajo FEDEA 95-20.

Garrido, L. (2012): «Estudios y trabajo de los españoles en la crisis de empleo», *Estudios de Economía Aplicada*, vol. 30-1, 29-58.

Iglesias, C.; Llorente, R., and Dueñas, D. (2010): «ICT Employment, Over-education and Gender in Spain. Do Information and communication Technologies Improve the Female Labor Situation?», *New Technology, Work and Employment* (Wiley-Blackwell), vol. 25 Issue 3, 238-252.

Motellón, E.; López-Bazo, E., and El-Attar, M. (2011): «Regional Heterogeneity in Wage Distributions: Evidence from Spain», *Journal of Regional Science*, vol. 51 (3), 558-584.

Rahona, M. (2008): «Un análisis del desajuste educativo en el primer empleo de los jóvenes», *Revista Principios-Estudios de Economía Política*, 11, 45-67.

Rahona, M.; Murillo, I., and Salinas, M. (2010): «Incidencia del desajuste educativo en el rendimiento privado de la educación en España», *Investigaciones de Economía de la Educación,* vol. 5, chapter 13, 267-284.

Serrano, L. (2002): «Salarios regionales y dotaciones de capital humano», *Revista de Economía Aplicada*, 10 (28), 23-38.