Human Capital and Regional Wage Gaps

Journal:	*Regional Studies*
Manuscript ID:	CRES-2009-0334.R2
Manuscript Type:	Special Issue Paper
JEL codes:	C24 - Truncated and Censored Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, J31 - Wage Level, Structure; Differentials by Skill, Occupation, etc. < J3 - Wages, Compensation, and Labor Costs < J - Labor and Demographic Economics, R11 - Regional Economic Activity: Growth, Development, and Changes < R1 - General Regional Economics < R - Urban, Rural, and Regional Economics, R23 - Regional Migration|Regional Labor Markets|Population < R2 - Household Analysis < R - Urban, Rural, and Regional Economics
Keywords:	Education and Experience, Regional disparities, Returns to human capital, Wage gap decomposition

Human Capital and Regional Wage Gaps

Enrique López-Bazo♂

Elisabet Motellón♀

♂ European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS). C/ Inca Garcilaso 3, E–41092, Seville, Spain
&
AQR–IREA, University of Barcelona, Avda. Diagonal 690, E–08034, Barcelona, Spain
Tel: +34 93 4037041 FAX: +34 93 4021821 Email: elopez@ub.edu

♀ Universitat Oberta de Catalunya
Av. Tibidabo 39-43, E-08035, Barcelona, Spain
Tel: + 34 932537501 FAX: +34 934176495 Email:emotellon@uoc.edu
&
AQR—IREA, Universitat de Barcelona

(Received October 2009 : in revised form April 2011)

ABSTRACT

This paper uses micro-level data to analyse the effect of human capital on regional wage differentials. The results for the set of Spanish regions confirm that they differ in the endowment of human capital, but also that the return that individuals obtain from it varies sharply across regions. Regional heterogeneity in returns is especially intense in the case of education, particularly when considering its effect on the employability of individuals. These differences in endowment and, especially, in returns to human capital, account for a significant proportion of regional wage gaps.

Keywords: Education, Experience, Regional disparities, Returns to human capital, Wage gap decomposition

JEL: C24, J31, R11, R23

1. INTRODUCTION

The effect of human capital – an intangible asset embodied in individuals – on regional growth and development has been examined by regional scientists and economists in recent decades. The assumption has been that the human capital endowment of a regional economy is an essential element in explaining its level of development and long-run economic growth. Besides its effect as an additional factor of production, it has been argued that human capital allows and encourages the generation and adoption of technological innovations that improve productivity. Almost all the studies to date have used aggregate data for a set of regions, and so the key variables considered have been the average of the measure used to proxy for the endowment of human capital (e.g. average years of schooling or the share of population with a certain educational attainment) in each region and some measure of aggregate economic activity, such as income or output per capita. In addition, previous studies have only considered the possibility that regional differences in levels of development and growth are due to different human capital endowments across regions (RODRÍGUEZ-POSE and VILALTA-BUFÍ, 2005; DI LIBERTO, 2008; LÓPEZ-BAZO and MORENO, 2008; BRONZINI and PISELLI, 2009). That is, no attention has been paid to the possibility that regional heterogeneity in the impact of human capital may be the cause and the effect of some of the economic disparities observed across regions. However, KRUEGER and LINDAHL (2001) showed that the effect of education on economic growth varies across countries. And even though it can be argued that regions within countries are likely to be more homogenous, our belief is that the assumption of equality of regional returns should be proved instead of imposed. Otherwise, the estimate of the average return in the sample of regions is likely to be biased. Accordingly, this paper argues that regions may differ in both the endowment and the return to human capital accumulated by individuals. Therefore, both should be considered when explaining regional differences in levels of economic activity.

To complement the evidence obtained by using aggregate data, this paper proposes the use of micro-data at the regional level. Micro-data provide additional evidence on the effect of human capital in explaining regional disparities and, in turn, a more appropriate control of regional differences in the distribution of individuals' characteristics. In particular, the use of individual data makes it possible to quantify the degree of regional differences in human capital endowment and also to measure its specific effect in each region, that is, to check whether the regions are also heterogeneous in the returns they obtain from human capital investments made by individuals. This has obvious implications for assessing policies designed to increase human capital endowment in order to promote growth in the less developed regions, as the effectiveness of such policies largely depends on the particular effect that human capital has in each region.

The use of information at the individual level allows a consideration of two different effects of human capital on regional economic performance. The first is the immediate effect on productivity from those in employment. The second is an indirect effect that is likely to occur through the increased employability of individuals endowed with a certain level of human capital. Studies using aggregate regional data have focused only on the first of these effects, although there is evidence to support a positive effect of human capital on labour market participation and a negative influence on the likelihood and duration of episodes of unemployment. Our hypothesis is that the two types of effect may differ across regions, thus contributing to regional disparities.

Using reliable individual data on wages obtained from a representative survey for each Spanish region, this paper assesses the effect of human capital within the framework of a Mincerian wage equation. In doing so, it follows DE LA FUENTE et al. (2003) and CICCONE et al. (2004) that have recently applied a similar approach for the set of Spanish and Italian regions respectively. Under the human capital theory (BECKER, 1964), the higher a worker's human capital endowment, the higher the wage she will earn, since it is assumed that education and experience (the two traditional components of individuals' human capital) have a positive effect on her productivity.[1] However, these previous studies only provide estimates on the regional returns to schooling, and not on the contribution of regional differences in endowments and returns to the explanation of regional wage gaps. As a novel and major contribution, in this paper we analyse the role played by human capital in explaining regional wage gaps, the hypotheses being that i) in addition to the effect associated with regional differences in human capital endowment, heterogeneity in terms of its return across regions may play a key role in explaining regional wage gaps, and ii) there is a direct effect of human capital, since it affects productivity of employees, and an indirect effect, by increasing the employability of all individuals. Aggregating over the individuals in a given region, this means that human capital stimulates aggregate productivity and the employment rate, thus contributing to increasing regional income per capita.

From a methodological point of view, the paper provides a framework for assessing regional differences in the conditional (being in employment) and the unconditional returns to education and experience. In a second step, it proposes a detailed decomposition of regional wage gaps to isolate the particular contribution of individuals' human capital. The approach followed here has been common practice for decomposing wage gaps across different groups of workers (e.g. gender or racial gap) in the labour market literature. But its application to the analysis of wage differentials across regions has been limited so far, and constrained to models that do not control for individuals' decision to participate in the labour market (REILLY, 1992; GARCÍA and MOLINA, 2002). The results for the set of Spanish regions confirm differences in terms of human capital endowment, and also in the return that individuals obtain in each region. Regional heterogeneity in returns is especially intense in the case of education, particularly when they incorporate the indirect effect. The decomposition of the wage gap between each region and the rest of the country shows that these differences in the endowment and in the returns to human capital account for a significant portion of the gap.

However, it should be kept in mind that micro-level analyses are less well suited for uncovering the social or aggregate return to human capital than macro growth studies. KRUEGER and LINDAHL (2001) showed that the differenced macro-Mincer equation, that results from aggregating the standard wage specification used to estimate the returns to human capital, differs from the typical macro growth model in several respects. For that reason, this paper just focuses on the contribution of the endowment and the return to human capital in explaining differences across regions in the wages received by individuals. Still, and given the connection between wages and productivity, it is our belief that something can be inferred from our results on the contribution of human capital in explaining differences across regions in economic growth and level of economic activity. But this should be made with caution considering the specific problems involved in the analysis of the growth effects of human capital at the aggregate level.

The rest of the paper is organized as follows. The next section briefly summarises the previous literature on the study of regional wage gaps. Section 3 introduces the dataset and discusses the results of the descriptive analysis. The empirical wage model and the derivation of the returns to the components of human capital are sketched in Section 4, which also discusses the results obtained for the set of Spanish regions. Section 5 presents the method proposed to obtain the detailed decomposition of the regional wage gaps and discusses the results of the contribution of human capital. Finally, section 6 concludes.

2. BRIEF LITERATURE REVIEW ON REGIONAL WAGE GAPS

A number of studies in the last three decades have aimed at quantifying the magnitude of regional wage gaps and identifying their origin. Most of these empirical analyses have been guided by two classical ideas: the fact that regional labour markets are heterogeneous and that there are compensating differentials that offset differences in price levels and non pecuniary attributes across regions. As a result, the real wage paid to each class of worker should be interregionally invariant. In other words, the competitive model behind these assumptions suggests that the price of the characteristics that determine wages will converge across regions in the absence of imperfect information and persistent stochastic disturbances, and with some mobile factors (see for instance FARBER and NEWMAN, 1989).

Apart from those derived from competitive theories, other sources of persistent regional gaps in real wages have been suggested. BLACKABY and MURPHY (1991) identify the role of labour market institutions (unionisation and the bargaining system), the determination of an individual's reservation wage, and variants of efficiency wage theory as possible explanations for the observed persistence of regional wage premiums. In the case of efficiency wage models, FARBER and NEWMAN (1989) argue that interregional differences in the conditions requiring efficiency wage premiums (turnover, shirking, adverse selection, threats of unionisation, worker's morale, etc.) will cause persistent differences across regions for identical workers.

A first bunch of studies (e.g. GERKING and WEIRICK, 1983; DICKIE and GERKING, 1987) tested for interregional structural shifts in the wage equations estimated, and concluded that real wages did not differ between macro-regions in the U.S. Rather, differences in average real wages arose from heterogeneous worker characteristics. In sharp contrast, FARBER and NEWMAN (1987) followed an alternative empirical approach that produced quite different result. They applied a decomposition method that allowed them to determine the contribution of differences in returns and endowments in accounting for wage differences across U.S. regions. Their results suggest that differences in returns (prices) may be at least as important in accounting for regional wage differentials as differences in worker characteristics between regions. Similar conclusions were obtained for the U.K. and other E.U. Member States using analogous decompositions of the regional wage gaps (MAIER and WEISS, 1986; BLACKABY and MANNING, 1990; REILLY, 1992; BLACKABY and MURPHY, 1995; GARCÍA and MOLINA, 2002).

In the particular case of Spain, GARCÍA and MOLINA (2002) analysed the wage gaps between the NUTS III Spanish regions using the 1994 wave of the European Community Household Panel (ECHP). They conclude that the contribution of differences in the return to human capital was only marginal. However, it should be stressed that there are some important differences between their analysis and the one in this paper. Firstly, the regional breakdown they used in their analysis is an artificial one that results from an arbitrary geographical grouping of the NUTS II Spanish regions (corresponding to the 17 Autonomous Communities, which are historical geographical and administrative regions with a high level of political and financial autonomy). The use of the ECHP wave for 2000 allows us to deal directly with the NUTS II Spanish regions. Secondly, as they did not account for the endogeneity of schooling, their estimates of the regional wage equations used to compute the wage decompositions are likely to be biased. Finally, they only considered the direct effect of human capital on wages, thus neglecting the effect of schooling on the probability of being employed. Given the high rates of unemployment and low participation, especially in some regions, it is reasonable to assume that the impact of education on employment was relevant in Spain.

3. DATASET AND DESCRIPTIVE ANALYSIS

This paper uses the micro-data from the Spanish sample of the ECHP.[2] The ECHP is a standardized survey conducted in the Member States of the European Union under the auspices of the Statistical Office of the European Communities (EUROSTAT). The survey involved annual interviewing of a representative panel of households and individuals in each country. The analysis in this paper exploits the 2000 extended sample of the ECPH because it was specifically designed for cross-sectional studies and above all because it is the only wave that provides representative samples at the NUTS II regional level in Spain. NUTS is the French acronym for Nomenclature of Territorial Units for Statistics, a hierarchical classification established by EUROSTAT which provides comparable regional breakdowns of EU Member States. In Spain, the NUTS II regions correspond to the 17 Autonomous Communities, which are historical geographical and administrative regions with a high level of political and financial autonomy.[3] The ECHP offers detailed information on the personal characteristics of the individuals and on the household, as well as on the labour conditions of those employed.

For the analysis of the effect of human capital on regional wages, the sample of individuals between 16 and 65 years in all the Spanish regions, except for the two city-regions in the north of Africa (Ceuta and Melilla), has been selected. Specifically, in the case of employees our sample includes full time non self-employed workers, defined in the ECPH as those working 30 or more hours a week, and those that, despite working less than 30 hours, declare themselves in the survey as full time workers.[4]

A first insight into the amount of regional wage differentials in Spain is obtained from the simple description of the sample in Table 1, which in the first two columns of data shows the average and the standard deviation of the gross hourly wage, and the number of workers contained in the sample for each one of the regions and for Spain as a whole. Large differences in average wages across regions are observed. For instance, the average wage in Extremadura, the region with the lowest wage level, was only 69.75% of the average wage in the Basque Country, the region with the highest. And the ratio between the top five regions and the five bottom regions is 1.29. This evidence confirms that the amount of regional wage disparities is of the same order of magnitude as those existing in other key economic variables such as income per capita and labour productivity.

<<<< INSERT TABLE 1 ABOUT HERE >>>>

In order to control for the effect of price differentials, an estimate of the relative level of regional prices has been used to compute real wages in each region. The regional relative price indexes are based on a representative basket that includes expenditures made by families in all goods and services (including housing) but that, however, imposes the same basket in all regions and time periods. Still, as far as we know this is the only information on relative regional prices available in Spain.[5] The average and standard deviation of real wages are shown in the third column of data in Table 1. Taking account of price differentials causes some changes in the ranking of regions, the most significant case being Extremadura, which moves from bottom to eighth place. Additionally, wage differentials are somewhat lower in real terms. For instance, the average real wage in the five bottom regions increases by around 2% due to their lower relative prices, whereas the average in the five upper regions falls by the same percentage as a result of their higher prices. However, most of the regional disparities remain after controlling for differences in prices across regions: for instance, the average real wage in Murcia (the region with the lowest value) is still under 75% of the real wage in Madrid, the region with the highest level in real terms.

Real wages may differ between regions because of what is known as the composition effect, that is to say, because workers' characteristics differ across regions.[6] In this case, the real wage paid to each class of workers should be interregionally invariant, and wage differentials would be merely an illusion caused by the failure to distinguish between types of labour (FARBER and NEWMAN, 1989). A simple look at the amount of regional differences in workers' characteristics in the sample can be obtained from Table 2, which shows the average value for the characteristics observed in the sample for the whole of Spain, and for the two regions with the highest (Madrid) and lowest (Murcia) average real wage.[7] In each case, the figures refer both to the sample of employees and non-employees (unemployed workers and non-participants). Focusing on the measures of human capital, the results reveal notable differences in education (measured by years of schooling) and in tenure between the two regions. On average, employees in Madrid spent more than two years longer at school than those in Murcia; the difference is not so high among non-employees, but it remains non-negligible (about 1 year). As regards tenure, most of the differences correspond to the categories of less than one year and more than 15 years. This is to do with regional differences in the number of fixed-term contracts; which is much higher in Murcia than in

Madrid (for a further discussion of this issue, see MOTELLÓN, 2008). In contrast, there do not seem to be significant differences across regions in labour market experience.

<<<< INSERT TABLE 2 ABOUT HERE >>>>

The last set of columns in Table 1 confirms that marked differences exist across all the Spanish regions in the variables proxying for the individuals' human capital, particularly in schooling and tenure. Interestingly, the comparison of the regional figures on real gross hourly wages with those of the human variables reveals a significant positive association between them. The linear correlation coefficient for regional averages of wages and schooling is as high as 0.80, while for wages and tenure (at least 10 years in the firm, that is the sum of the shares in the last two categories of tenure in Table 1) is 0.62. The association in the case of experience is not as clear, with a correlation coefficient of 0.34.

Table 2 shows differences between regions for other individual and household characteristics, such as gender, age and household composition, for both employees and non-employees. Therefore, wages may differ across regions because regions have different human capital endowments and because of other characteristics that are believed to affect wages directly and indirectly, through the probability of employment, but also because of regional differences in the return to human capital and in the price of other characteristics.

This seems to be supported by the wage differences observed within categories of levels of schooling, tenure and experience, as reported in Table 3. This table shows the average real wage for the sample of workers in each of the categories of the human capital variables, for Spain as a whole and for the regions with the highest and lowest average real wages. Observe that the wage gap between Madrid and Murcia at each level of schooling decreases somewhat, although the average wage in Murcia was still some 20% lower. The only exception is the regional gap for workers with a university degree, in which case the average wage in Murcia was 92% of that in Madrid. The cases of tenure and experience are quite similar, as the regional gap within categories decreases only marginally (the wage in Murcia being between 70% and 80% of that in Madrid for most of the categories).

<<<< INSERT TABLE 3 ABOUT HERE >>>>

Taking this preliminary descriptive evidence into consideration, our hypothesis is that not only the endowments but also the returns to human capital vary across regions, thus contributing to wage differentials, both directly and indirectly through the impact that human capital has on the probability of employment. The next section presents results for the estimates of direct and indirect effects of human capital obtained when conditioning to other factors that are also likely to affect the wage earned by each worker. The estimates of the returns to schooling, tenure and experience obtained for each Spanish region will allow us to check for the regional heterogeneity in the returns to human capital.

4. REGIONAL RETURNS TO HUMAN CAPITAL

4.1. Empirical framework

The framework for the empirical analysis is a model in which the wage for an individual i in region r is given by:

$$W_{ir} = X_{ir}\beta_r + \varepsilon_{ir} \quad (1)$$

$$C^*_{ir} = Z_{ir}\gamma_r + \upsilon_{ir} \quad (2)$$

where W_{ir} is the log of the wage of individual i in region r, X_{ir} denotes the set of characteristics that affect the wage of this individual in a direct way (education, experience and its square, tenure, and gender), and β_r is the vector of prices or returns associated with the characteristics.[8] C^*_{ir} is a latent and unobservable process that assigns the individual i in region r to the sample of employees or to the sample of non-employees, Z_{ir} being the vector of observations for characteristics that determine the process of selection (education, gender, age, marital status, chronic disease, proxies for household composition, and household income other than the wage of the individual)[9] and γ_r the corresponding parameters. ε_{ir} and υ_{ir} are i.i.d errors following a bivariate normal distribution $(0, 0, \sigma_{\varepsilon_r}, \sigma_{\upsilon_r}, \rho_r)$, with ρ_r the correlation coefficient for both error terms in region r.

Only the result of the selection process in (2) is observed, the indicator variable C_{ir}, that equals 1 when $C^*_{ir}>0$, and 0 otherwise. Then, the probability of employment (selection) of individual i in region r is given by:

$$C_{ir} = \text{Prob}(C^*_{ir} > 0) = \text{Prob}(\upsilon_{ir} > -Z_{ir}\gamma_r) = \Phi(Z_{ir}\gamma_r) \qquad (3)$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Estimates of returns based on the wage equation in (1), leaving aside the selection equation in (2), are biased and inconsistent if $\rho_r \neq 0$. Consistent estimates can be obtained by maximum likelihood considering the information from the two equations or, alternatively, by applying the two-step method proposed in HECKMAN (1979). The Heckit method includes the inverse Mills ratio in the wage equation as an additional regressor to obtain wages conditional on being employed:

$$W_{ir} \mid C^*_{ir} > 0 = X_{ir}\beta_r + \theta_r\lambda_{ir} + \varepsilon_{ir} \qquad (4)$$

where

$$\lambda_{ir} = \frac{\phi(Z_{ir}\gamma_r)}{\Phi(Z_{ir}\gamma_r)} \qquad (5)$$

is the inverse Mills ratio for individual i in region r computed from the probabilistic model in (3), and $\theta_r = \rho_r\sigma_{\varepsilon_r}$ is the coefficient that measures its effect on wages.[10]

From the specification of the model of wage determination in (1) and (2), and the one for conditional wages in (4), different types of returns to characteristics can be defined.[11] In the case of education – S – the conditional return is defined as:

$$CRS_{ir} \equiv \partial E\left[W_{ir} \mid C^{*}_{ir} > 0\right] \big/ \partial S_{ir} = \beta^{S}_{r} - \theta_{r}\gamma^{S}_{r}\delta_{i} \qquad (6)$$

where $E\left[W_{ir} \mid C^{*}_{ir} > 0\right] = X_{ir}\beta_{r} + \theta_{r}\lambda_{ir}$ and $\delta_{i} = \left(Z_{ir}\gamma_{r} + \lambda_{ir}\right)\lambda_{ir}$. Then, CRS_{ir} is the marginal effect of S_{ir} on the conditional expected value of W_{ir}. The second term is the correction that takes into account that only the effect of S_{ir} on W_{ir} for employed individuals should be considered. That is to say, CRS_{ir} is a measure of the effect that a year of education has on the wage received by employees. Notice that the conditional return to education will be different for each individual in each region, as it depends on the regional coefficients β^{S}_{r}, θ_{r}, and γ_{r}, and on the value of δ_{i}. As is usual in these cases, the conditional return to education for each region r – CRS_{r} – will be computed as the average for the sample of employees in that region.

In addition, the expected value of the wage earned by a randomly selected individual from the entire population (employees and non-employees) is of interest as well:

$$E\left[w_{ir}\right] = \Phi\left(Z_{ir}\gamma_{r}\right)E\left[w_{ir} \mid C^{*}_{ir} > 0\right] = \Phi\left(Z_{ir}\gamma_{r}\right)\exp\left(X_{ir}\beta_{r} + \theta_{r}\lambda_{ir} + 0.5\sigma^{2}_{\varepsilon_{r}}\right) \qquad (7)$$

where w_{ir} is the wage level of individual i in region r. That is, for any individual the unconditional expected wage is the one obtained in the case of being employed, multiplied by the probability of being employed. The marginal effect of education on the unconditional expectation in (7) is then defined as the unconditional return to education (provided that the function is evaluated at a point with $E\left[w_{ir}\right] \neq 0$):

$$URS_{ir} \equiv \frac{\partial E[w_{ir}]/E[w_{ir}]}{\partial S_{ir}} = \frac{\partial \ln E[w_{ir}]}{\partial S_{ir}} = \beta_r^S - \theta_r \gamma_r^S \delta_i + \gamma_r^S \lambda_i = CRS_{ir} + \gamma_r^S \lambda_i \qquad (8)$$

The second term in the unconditional return in (8) reflects the effect that education has on the probability of employment, which is an indirect effect on wages. As this effect is likely to be positive (more education will decrease the episodes of unemployment and non-participation), the URS_{ir} is expected to be higher than the CRS_{ir}. As stressed in ARRAZOLA and DE HEVIA (2008) individuals take this indirect effect into account when they decide on their investment in education. As in the case of the conditional return, URS_{ir} depends on regional coefficients and on individual values for the characteristics that determine the process of participation, Z_{ir}. Accordingly, the unconditional return to education for each region r – URS_r – will be computed as the average for the total sample of individuals (employees and non-employees) in that region.

As for the other two components of human capital, experience and tenure, note that they are not included in the list of determinants of the probability of employment. As a consequence, they only exert a direct influence on wages through their inclusion in the wage equation. This means that the unconditional effects of these characteristics equal the conditional ones, which are simply a function of the corresponding elements in the vector of coefficients of the wage equation, β.[12]

4.2. Results

The conditional and unconditional returns defined above were computed based on the estimation of the coefficients in the empirical wage model defined by (1) and (2). As already indicated, a simple specification for the wage equation was used to fully account for the effects of the human capital variables. It includes the number of years of schooling, the years of experience and its square, a set of dummies that account for tenure, and the gender of the individual. The dependent variable is the real gross hourly wage for the individuals in each region. As for the participation equation in (2), in addition to the measure of education, it includes proxies for the individual and family characteristics that are supposed to affect the chance of being employed: the individual's gender, age, and marital status, presence of chronic disease, the household income other than the wage earned by the individual, and variables of household composition such as its size, the number of children under 15 years, and the presence of children under 6 years.

An instrumental variables estimator (IV) was used to avoid the bias of the OLS estimates due to the likely endogeneity of education (see for instance CARD, 1999 and 2001).[13] Suitable instruments should capture exogenous factors that affect the choice of the individuals' degree of education but not their current wages. Immediate information on variables of this kind (such as family background and ability) is not readily available from surveys like the one used in this study. So we follow the suggestion made in the recent related literature and use as instruments variables that reflect whether the education of the individual was affected by profound changes in the educational system and by extraordinary historical events such as a war (see for instance HARMON and WALKER, 1995; ICHINO and WINTER-EBMER, 1999 and 2004; ARRAZOLA et al., 2003). Specifically, a dummy variable was defined to account for the effect of the change in the regulation of the Spanish educational system brought in by the 1970 General Education Act, which established free, compulsory education for children between 6 and 14 years old. The instrument is a dummy variable that takes a value of 1 for individuals aged 6 or under in 1971, that is, members of the sample whose period of schooling was affected by the reform. An instrument related to the Spanish Civil

War (lasting from 1936 to 1939) was also defined. In this case the aim is to capture the effects that the war and the post-war periods had on the education of individuals who were in schooling age during those years. Correspondingly, as in ARRAZOLA et al. (2003) and ARRAZOLA and DE HEVIA (2008) we defined a dummy variable taking a value of 1 for individuals born in or before 1945.[14] As argued by ICHINO and WINTER-EBMER (1999) this event would have affected the opportunities of education of different types of individuals, being the ones more affected those with high ability and facing liquidity constraints (i.e. those in poor families).

In addition, following the suggestion in WOOLDRIDGE (2002), the variables in Z, that is, the ones that affect the probability of employment, were included in the list of potential instruments for education in the wage equation. Taking into account the risk in terms of bias and inefficiency associated to the inclusion of invalid and/or weak instruments (see for instance MURRAY, 2006) we have chosen not to include the full list of variables in Z but to select the subset most appropriate for each region. The final set of instruments from Z was selected in each case based on the results of the battery of tests designed to check for the validity of instruments (under and weak identification, and exogeneity). Finally, it should be mentioned that the variable for the change in the educational system, the one for the Spanish Civil War, and the corresponding subset from Z were also used as instruments for the inverse Mills ratio, as it is a function of schooling and, hence, likely to be endogenous in the wage equation.[15]

Based on the existence of incomplete information about the worker's ability and/or the quality of the worker-firm match, and their likely correlation with experience and tenure, it can also be argued the endogeneity of these two variables (see for instance DUSTMANN and

MEGHIR, 2005 and WILLIAMS, 2009). The methods proposed to address such an issue involve data on changes in wages and job displacements, which are not available in the dataset we use in this study. Additionally, as far as we know those alternative estimation procedures are not compatible with the wage decomposition methods developed so far, including the one proposed and applied in this paper. Accordingly, we take experience and tenure as exogenous although we will interpret with caution the estimate of their returns and their contribution to the explanation of regional wage gaps.[16]

IV estimates for the parameters of the wage system in (1) and (2) –IV HECKIT– were obtained for each region and for Spain as a whole, jointly with the set of statistics for the validity of instruments.[17] The full set of results is not shown here for reasons of space, although some comments are in order. First, as shown in Table A.1 in the Appendix, the results of the tests of weak identification, underidentification and overidentifying restrictions suggest that, in almost all cases, the two variables accounting for the change in the educational system and the Civil War, and a subset of the variables in Z are appropriate instruments for schooling and the inverse Mills ratio. There is only one region (Asturias) in which the test proposed by STOCK and YOGO (2005) does not reject the null hypothesis of weak identification for a relative bias of the IV estimator lower than 30%. But still in this case, the underidentification test rejects the null hypothesis of no correlation between the set of instruments and the endogenous variables (meaning that the excluded instruments are *relevant*). Therefore, only for this region the estimate of the returns should be taken with some caution.

On the other hand, the coefficient for the inverse Mills ratio in the wage equation – θ_r – was significant, at 10%, in 8 out of the 17 regions and in the entire country (last column in Table

A.1). This means that for those cases the estimates based only on the wage equation, and thus ignoring the process of selection, would be biased. The estimate of θ_r was positive in 10 regions, and negative in 7 regions and in the entire country. This means that shocks that increase the probability of employment also increase the expected wage of employees in the former group of regions, while decrease it in the latter. Second, the coefficients in the wage and in the selection equations were jointly significant in all cases, particularly for the human capital variables. For all the regions, education increases the wage earned and the probability of receiving a wage. Experience and tenure also exert a significant positive effect on wages.[18]

Tables 4 and 5 reproduce the returns to the different types of human capital computed using the estimates above. As for the returns to education, the first two columns of results in Table 4 show the conditional and unconditional returns from the IV HECKIT estimates. Remember that the conditional return for each region was computed with the sample of employees, whereas the unconditional one was calculated with the whole sample (employees, unemployed and non-participants). Both types of returns were statistically significant at 1% in Spain and in all regions (for this reason, asterisks are not included alongside the figures in Table 4). The conditional return to education in the entire country was 7.85%, which means that an additional year of education increased the expected wage of those actually earning a wage by almost 8%. But this figure for the country as a whole hides significant regional heterogeneity in the conditional effect of education. The conditional return in Extremadura and Murcia is far above that in Asturias. Among the regions with the highest conditional return are some of the traditionally less developed regions, which are also among the regions with the lowest endowments of education (see Table 1). In contrast, the return was below the country average in some of the most advanced regions, which are the ones with the highest endowment of that type of human capital (such as Madrid, the Basque Country, and Aragon).

<<<< INSERT TABLE 4 ABOUT HERE >>>>

As for the unconditional return to education, the second column of results in Table 4 shows that in Spain as a whole it was far above the conditional return. An increase of one year of schooling represented an increase of more than 18% in the expected wage of an individual randomly drawn from the Spanish active population. This result confirms the importance of considering the indirect effect of education when analysing its connection with wage expectations. Actually, the estimate for Spain suggests that the second term of the unconditional return defined in (8) – the indirect effect – is far larger than the direct effect of education on employees' productivity. The high and persistent unemployment rates in Spain, and the already low figures on participation in the labour market (especially for women), jointly with the strong effect of schooling on the likelihood of unemployment and on participation decisions, explains this result. Accordingly, our intuition is that this indirect effect of education is larger in Spain than in other developed economies with different labour market outcomes. The same argument applies to all the regions under analysis, although once again the results for the estimates of the unconditional returns at the regional level confirm our hypothesis of the strong spatial heterogeneity in the effect of education. The unconditional return in Asturias (11.1%) is almost a third that in Extremadura, which is as high as 29.2%. And regardless of some changes in the ranking, the association between returns and the level of development (and in this case of employment rates) is also observed for unconditional returns.

Table 4 also includes the returns to education computed from the OLS estimates (considering schooling and the inverse Mills ratio as exogenous, and no correlation between the error terms in equations (1) and (2), that is $\rho_r=0$), the IV estimates (assuming that $\rho_r=0$ but not imposing the exogeneity of schooling and the inverse Mills ratio), and the HECKIT estimator (imposing the exogeneity of schooling and the inverse Mills ratio but allowing for a selection process). It can be observed how, in agreement with the previous results in the literature, the IV estimate of the returns to schooling is as big or even bigger than the one from the least square estimates. Differences are also observed when comparing results from estimates controlling for sample selection, although in that case they are lower in magnitude.

All in all, these estimates confirm the positive (direct and indirect) effect of education on wages and the existence of substantial regional variability in the return to investments in this type of human capital. In addition, the results in Table 5 show that there was also regional heterogeneity in the return to the other types of human capital considered in this study: general experience in the labour market and specific experience in the firm (tenure). In the country as a whole, an additional year of general experience caused an increase of around 1.2% in the expected wage. The return to experience is much higher in regions such as Extremadura and Galicia (1.93% and 1.81% respectively) and substantially below the country average in others like the Balearic Islands (0.74%), Asturias (0.77%), and Cantabria (0.78%). The case of returns to tenure is quite similar, as the profile of wage increases associated with the defined intervals of years of specific experience varies widely across regions. For instance, in some regions (Asturias, Aragon, Canary Islands, and Madrid) there was a substantial gain linked to workers' tenure: employees with 15 and more years' experience in the firm earned above 30% more than those with one or less than one year. This gain was far lower in some other regions (Andalusia, Balearic Islands, Cantabria, and Galicia), and non-significantly different from zero in Extremadura. As in the case of schooling, the return to experience and tenure was also computed based on the OLS, IV and HECKIT estimators. The pattern observed in these cases is quite similar to the one from the IV HECKIT estimates, particularly as regards the regional heterogeneity in returns.[19]

<<<< INSERT TABLE 5 ABOUT HERE >>>>

The evidence presented so far thus not only confirms that regions differed in the human capital endowment of their employees and the rest of their labour force but also shows sizeable regional variability in the return that individuals obtain from their accumulated human capital. As the final step in this study, the next section assesses the contribution of this variability in regional endowments and returns to the wage gap across regions.

5. HUMAN CAPITAL AND REGIONAL WAGE GAPS

5.1. Methodology

This section briefly describes the method proposed to obtain a detailed decomposition of the average wage gap between any two regions (A and B), or between a region and the rest of the country, under the presence of a selection process such as the one described in (2). Technical details of the derivation are sketched in the appendix. From expression (4), the average of conditional (log) wages in regions A and B can be expressed as:

$$\overline{W}_A = \overline{X}_A \hat{\beta}_A + \hat{\theta}_A \overline{\lambda}_A \qquad (9)$$

$$\overline{W}_B = \overline{X}_B \hat{\beta}_B + \hat{\theta}_B \overline{\lambda}_B \qquad (10)$$

where the "over bar" represents the value of the sample's average. Defining the average of a counterfactual inverse Mills ratio for region B as:

$$\overline{\lambda}_B^A \equiv \overline{\frac{\varphi(Z_B \hat{\gamma}_A)}{\Phi(Z_B \hat{\gamma}_A)}} \qquad (11)$$

the difference between the second terms in the RHS of equations (9) and (10) can be expressed as:

$$\left(\hat{\theta}_A \overline{\lambda}_A - \hat{\theta}_B \overline{\lambda}_B\right) = \hat{\theta}_A\left(\overline{\lambda}_B^A - \overline{\lambda}_B\right) + \hat{\theta}_A\left(\overline{\lambda}_A - \overline{\lambda}_B^A\right) + \left(\hat{\theta}_A - \hat{\theta}_B\right)\overline{\lambda}_B \qquad (12)$$

Building on (12), NEUMAN and OAXACA (2004) proposed an extension of the traditional decomposition as follows: [20]

$$\overline{W}_A - \overline{W}_B = \left(\overline{X}_A - \overline{X}_B\right)\hat{\beta}_A + \hat{\theta}_A\left(\overline{\lambda}_A - \overline{\lambda}_B^A\right) + \overline{X}_B\left(\hat{\beta}_A - \hat{\beta}_B\right) + \hat{\theta}_A\left(\overline{\lambda}_B^A - \overline{\lambda}_B\right) + \left(\hat{\theta}_A - \hat{\theta}_B\right)\overline{\lambda}_B \quad (13)$$

The first two terms in the RHS of (13), $\left(\overline{X}_A - \overline{X}_B\right)\hat{\beta}_A + \hat{\theta}_A\left(\overline{\lambda}_A - \overline{\lambda}_B^A\right)$, correspond to differences in the endowment of characteristics between regions A and B, both those directly affecting wages and those determining the probability of employment. In other words, that would have been the wage gap between regions A and B if they had differed only in the endowment of characteristics. The third and fourth terms, $\overline{X}_B\left(\hat{\beta}_A - \hat{\beta}_B\right) + \hat{\theta}_A\left(\overline{\lambda}_B^A - \overline{\lambda}_B\right)$, measure the contribution to the wage gap of regional heterogeneity in returns, through the direct and the indirect effect respectively. It is interpreted as the gap we would have observed if regions A and B had differed in the return to the characteristics only. Finally, $\left(\hat{\theta}_A - \hat{\theta}_B\right)\overline{\lambda}_B$ is a sort of residual term related to the regional difference in the impact of the process of selection on wages.

The decomposition in (13) allows us to assess the contribution of characteristics and returns to the regional wage gap including the indirect effect coming from the process of selection. Therefore it is a decomposition of the gap in conditional wages. However, it does not allow us to obtain the contribution of each characteristic and each group of characteristics. This would be of particular interest when, as in this paper, we are interested in the effect of a set of variables such as those proxying for workers' human capital. The problem is how to assign the individual contribution to each variable when a non-linear term is involved; the actual and counterfactual inverse Mills ratios in equation (13). Our proposal to overcome this problem builds on YUN (2004)'s general decomposition of gaps in the first moments when the variable under analysis depends on a non-linear function which, however, has a linear function as argument. In this case, the decomposition in (13) can be expressed as:

(14)

$$\overline{W}_A - \overline{W}_B = \left\{ \sum_{i=1}^{l_X} P^i_{\Delta X} \left[\left(\overline{X}_A - \overline{X}_B \right) \hat{\beta}_A \right] + \sum_{i=1}^{l_Z} P^i_{\Delta Z} \left[\hat{\theta}_A \left(\overline{\lambda}_A - \overline{\lambda}^A_B \right) \right] \right\} +$$
$$\left\{ \sum_{i=1}^{l_X} P^i_{\Delta\beta} \left[\overline{X}_B \left(\hat{\beta}_A - \hat{\beta}_B \right) \right] + \sum_{i=1}^{l_Z} P^i_{\Delta\gamma} \left[\hat{\theta}_A \left(\overline{\lambda}^A_B - \overline{\lambda}_B \right) \right] \right\} +$$
$$\left(\hat{\theta}_A - \hat{\theta}_B \right) \overline{\lambda}_B$$

where

$$P^i_{\Delta X} = \frac{\left(\overline{X}^i_A - \overline{X}^i_B \right) \hat{\beta}^i_A}{\left(\overline{X}_A - \overline{X}_B \right) \hat{\beta}_A}, \qquad P^i_{\Delta Z} = \frac{\left(\overline{Z}^i_A - \overline{Z}^i_B \right) \hat{\gamma}^i_A}{\left(\overline{Z}_A - \overline{Z}_B \right) \hat{\gamma}_A}$$

$$P^i_{\Delta\beta} = \frac{\overline{X}^i_B \left(\hat{\beta}^i_A - \hat{\beta}^i_B \right)}{\overline{X}_B \left(\hat{\beta}_A - \hat{\beta}_B \right)}, \qquad P^i_{\Delta\gamma} = \frac{\overline{Z}^i_B \left(\hat{\gamma}^i_A - \hat{\gamma}^i_B \right)}{\overline{Z}_B \left(\hat{\gamma}_A - \hat{\gamma}_B \right)}$$

$$\sum_{i=1}^{l_X} P^i_{\Delta X} = \sum_{i=1}^{l_Z} P^i_{\Delta Z} = \sum_{i=1}^{l_X} P^i_{\Delta\beta} = \sum_{i=1}^{l_Z} P^i_{\Delta\gamma} = 1$$

are the weights that allow us to assign the contribution of each variable in X and Z to differences in characteristics ($P^i_{\Delta X}$ and $P^i_{\Delta Z}$) and in returns ($P^i_{\Delta\beta}$ and $P^i_{\Delta\gamma}$).[21] l_X and l_Z denote the number of characteristics included in X and in Z respectively.

5.2. Results

Instead of decomposing the wage gap for each pair of the 17 Spanish NUTS II regions, we computed the global and the detailed decomposition for the gap between the rest of the country and each region r, that is $\left(\overline{W}_{SP-r} - \overline{W}_r \right)$, where $\overline{W}_{SP-r}$ is the average (log) real hourly gross wage for the sample of employees in Spain excepting those in region r, and $\overline{W}_r$ is the corresponding average for region r.[22] Then, following the notation in the previous section, A corresponds to SP-r, and B corresponds to r. To implement the decomposition of those gaps, we used the IV HECKIT estimates of the coefficients in the wage and in the selection equations, β and γ, for each region, which were described in section 4.2. A set of IV HECKIT estimates for the same coefficients was obtained corresponding to the samples of the rest of the country associated with each region. The characteristics of these estimates were similar to those discussed in section 4.2 in the case of the entire country.

As a first step, the results obtained for the global decomposition in equation (13) are summarized in the first set of columns in Table 6. The first column of results shows the regional wage gap as defined above. It is positive when the average wage in the rest of the country exceeds the average wage in the region, and negative when the wage is higher in the region. The second and third columns of results correspond to the contribution of differences in endowments and returns to all the characteristics as defined in equation (13). Finally, the fourth column contains the contribution of the residual of the process of selection which depends on the average of the inverse Mills ratio in the region, and on the difference between the coefficient θ in the region and in the rest of the country: that is, the part of the wage gap attributed to differences in the particular impact of the probability of employment on the wage level.

<<<< INSERT TABLE 6 ABOUT HERE >>>>

As a matter of example, the wage gap (in logs) between the rest of the country and Andalusia was 0.0342, or in other words, the average wage in Andalusia was 3.42% lower than the average wage in the rest of the country. If Andalusia and the rest of the country would have only differed in the endowment of all the characteristics (X and Z), the gap had been even higher (0.0648). This is the contribution to the actual gap of differences in endowments for that region. But the actual gap was lower because differences in returns favoured Andalusia: if the only difference would have been the one in returns, the average wage in that region had been slightly higher than in the rest of the country. That is the reason for the negative coefficient of the contribution of returns in that case (-0.0015). Finally, the residual associated to the process of selection also counteracts partially the effect of differences in endowments. Its contribution is estimated to be -0.0290, which means that $\hat{\theta}_{AND}$, the estimated impact of the probability of employment (through the inverse Mills ratio) on wages in Andalusia, is higher than that for the rest of the country.

A negative wage gap is obtained when the average wage in the region under analysis is higher than that in the rest of the country, as in the case of the Basque Country (-0.1666). In that case, both differences in endowments and in returns favoured that region, as the contribution of the two components is negative in both cases (-0.0815 and -0.0263). Finally, the residual of the selection process also contributed to a higher average wage in the Basque Country (-0.0588).

From the results of the global decomposition, it can be concluded that the contribution of returns is almost as large as that of endowment for most of the Spanish regions. Actually, it is particularly intense in regions with a positive gap. In these cases the contribution of returns clearly exceeds that of endowments (Asturias, Balearic Islands, Cantabria, Castile-La Mancha, Galicia, La Rioja) or both are of the same order of magnitude (Extremadura). Interestingly, in Asturias, Cantabria and in La Rioja, the contribution of returns was so favourable to the rest of the country that it counterbalanced the contribution of the other elements (endowment and residual term) that favoured those two regions. In sharp contrast, differences in endowments seem to explain most of the gap for regions with wages far above the rest of the country (Aragon, the Basque Country, Castile-Leon , and Madrid). Finally, it should be stressed that the contribution of the residual term in the decomposition in equation (13) is particularly intense for some regions, counterbalancing that associated with differences in returns in such cases (as in Castile-La Mancha, Murcia, and Valencia).

The specific contribution of human capital to the wage gap in each region has been computed using the detailed decomposition in equation (14). The effects of differences in endowment and in returns to human capital are shown in the fifth and sixth columns of results in Table 6. Taking again the case of Andalusia, results indicate that differences in the endowment of human capital alone would have caused wages in the rest of the country to be 7.90% higher than in that region. However, the contribution of differences in returns to human capital had the opposite effect, and it was larger in magnitude, counterbalancing the effect of the lower endowment of human capital in Andalusia. In any case, the results for this region illustrates quite well the prominent role of human capital in the explanation of the regional wage gaps, in particular that played by regional heterogeneity in its return. In fact, the results for the entire set of regions confirm that the effect of differences in returns is much larger than that of endowments in almost all cases.

It can also be deduced that the actual gap was much lower than the one that would have resulted from differences in human capital because differences in other characteristics partially counterbalanced the effect of this type of capital.

An interesting feature is observed in regions with wages below those in the rest of the country (positive gap). In Andalusia, Asturias, Castile-La Mancha, Extremadura, Galicia, and to a lesser extent in the Balearic Islands, the endowment of human capital contributed to the lower wages. But, in all cases, this effect was compensated by the large contribution of returns. This was not so, however, for some other regions with low wages in which both effects worked in the same direction such as the Canaries, Murcia and Valencia (where both endowments and returns to human capital contributed to widen the gap), and as Cantabria and La Rioja (where human capital contributed to narrow the gap). In any case, the results confirm that the contribution of differences in returns to human capital was greater than that of endowments for most of the regions.

The final step in our analysis was to isolate the particular contribution of schooling to regional wage gaps. In the descriptive analysis in section 3 and in the discussion of the estimated returns in section 4.2, it was observed that regional heterogeneity was more intense in education than in the two types of workers' experience. Correspondingly, we expected that most of the effect of human capital would come from the contribution of regional differences in education. The last two columns in Table 6 show the contribution of endowments and returns to education respectively. These figures confirm that most of the effects mentioned above in reference to human capital are related to education. For instance, in the case of Andalusia the contribution of differences in the endowment of education was 0.0542, which is more than two-thirds of the contribution observed for human capital endowments (0.0790). As for the contribution of differences in the return to schooling, it was half the total effect attributable to human capital (-0.0712 and -0.1401 respectively). Actually, the share of the effect attributable to schooling in the total contribution of human capital in the rest of the regions is even larger than in Andalusia. This is so both for differences in endowments and in returns.

All in all, the results in this section support our hypothesis regarding the role played by differences in endowment and also in returns to human capital to explain regional wage gaps. Similarly, within human capital, the crucial elements are the endowment of individuals' education and the return that they obtain from it.

6. CONCLUDING COMMENTS

The results of this study confirm the usefulness of using micro-data in studies dealing with regional economic disparities and the impact of intangible assets in each region. They provide complementary empirical evidence to that obtained from aggregated regional data. In the specific case of human capital, the use of individual data allowed us to evaluate both the impact of differences in the endowment and the return obtained by individuals within each region.

We show that there are significant regional differences in the distribution of education and experience in Spain. We also provide evidence of the existence of strong disparities in the return to human capital, especially in the case of education. Actually, the results suggest that a large proportion of the total effect of education is related to an indirect effect, since the impact of education on employability varies considerably from region to region. The detailed decomposition of the regional wage gaps has allowed us to demonstrate that regional heterogeneity in the returns to human capital was an important factor explaining wage disparities across regions. Moreover, the detailed results suggest that most of this effect should be attributed to differences in the return to education, since the differences associated to returns to tenure and experience played a minor role in most regions.

An immediate implication can be drawn from these results. It appears that policies aiming to promote education are an effective tool in improving workers' productivity and in lowering the risk of unemployment and non-participation in the labour market. The effect of these policies is also likely to be stronger in regions with lower levels of development. Therefore, raising educational attainment in these regions would contribute to regional convergence in labour productivity and unemployment and participation rates. The overall effect would thus be an increase in the average income per capita of the less favoured regions and a reduction in regional disparities. Also worth noting is the suggestion that the promotion of education in less developed regions simultaneously meets the goals of equity and efficiency, given that the return of this policy is higher in less developed regions than in more advanced ones.

It must be emphasized that the conclusions are derived from a partial equilibrium exercise. As is usual in exercises of this kind, the counterfactual analysis in this paper did not predict the reaction of workers and firms, for instance, to the regional equalization of endowments and/or returns to human capital. The counterfactual analysis has also considered the same selection process for participation and for employment, even though there are arguments to question this assumption. In any case, it should be stressed that ARRAZOLA and DE HEVIA (2008) did not report substantial differences in the estimate of the conditional and the unconditional returns to schooling for Spain as a whole, when a two-step selection process was used. And that the consideration of two different, and sequential, selection processes would have made more cumbersome the derivation of the detailed wage decomposition used in section 5. On the other hand, the system of collective bargaining existing in Spain may be inducing wage differences between regions, independently of workers' characteristics; differences in returns may then be related to differences in sectoral minimum wages determined at subnational level (SIMÓN et al., 2006). Still, our feeling is that the contribution of this element to the estimated regional differentials in the return to education is not as important as to invalidate the results discussed above. Actually, we are more sympathetic with alternative explanations for differences in returns, as the one derived from models of the New Economic Geography that relates incentives of individuals to invest in human capital to the market access of their locations (see for instance REDDING and SCHOTT, 2003, and COMBES et al., 2008). A deeper analysis of these points is on our future research agenda.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Spanish Ministry of Science and Technology, National Program of R&D, ECO2008-05314/ECON, and from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216813, IAREG Project. The views expressed here are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

APPENDIX

Evaluating the values of the inverse Mills ratios involved in the RHS of (12) using mean characteristics results in:

$$\left(\hat{\theta}_A\bar{\lambda}_A - \hat{\theta}_B\bar{\lambda}_B\right) = \hat{\theta}_A\left(\tilde{\lambda}_B^A - \tilde{\lambda}_B\right) + \hat{\theta}_A\left(\tilde{\lambda}_A - \tilde{\lambda}_B^A\right) + \left(\hat{\theta}_A - \hat{\theta}_B\right)\bar{\lambda}_B + R_M \qquad \text{(A.1)}$$

where $\tilde{\lambda}_r = \left(\varphi(\bar{Z}_r\hat{\gamma}_r)/\Phi(\bar{Z}_r\hat{\gamma}_r)\right)$, r=A, B, and $\tilde{\lambda}_B^A = \left(\varphi(\bar{Z}_B\hat{\gamma}_A)/\Phi(\bar{Z}_B\hat{\gamma}_A)\right)$. The error of approximation, R_M, is:

$$R_M = R_{MA} + R_{MB} + R_{MB}^A \qquad \text{(A.2)}$$

$$R_{Mr} = \bar{\lambda}_r - \tilde{\lambda}_r = \overline{\frac{\varphi(Z_r\hat{\gamma}_r)}{\Phi(Z_r\hat{\gamma}_r)}} - \frac{\varphi(\bar{Z}_r\hat{\gamma}_r)}{\Phi(\bar{Z}_r\hat{\gamma}_r)} \quad ; \quad r = A, B$$

$$R_{MB}^A = \bar{\lambda}_B^A - \tilde{\lambda}_B^A = \overline{\frac{\varphi(Z_B\hat{\gamma}_A)}{\Phi(Z_B\hat{\gamma}_A)}} - \frac{\varphi(\bar{Z}_B\hat{\gamma}_A)}{\Phi(\bar{Z}_B\hat{\gamma}_A)}$$

Using a first order Taylor expansion to linearize the terms that involve the inverse Mills ratios, $\hat{\theta}_A\left(\tilde{\lambda}_A - \tilde{\lambda}_B^A\right)$ and $\hat{\theta}_A\left(\tilde{\lambda}_B^A - \tilde{\lambda}_B\right)$, around $\bar{Z}_A\hat{\gamma}_A$ and $\bar{Z}_B\hat{\gamma}_B$ respectively:

$$\hat{\theta}_A\left(\hat{\lambda}_A - \hat{\lambda}_B^A\right) = \hat{\theta}_A\bar{f}_A\left(\bar{Z}_A - \bar{Z}_B\right)\hat{\gamma}_A + R_{T1}$$

$$\hat{\theta}_A\left(\hat{\lambda}_B^A - \hat{\lambda}_B\right) = \hat{\theta}_A\bar{f}_B\bar{Z}_B\left(\hat{\gamma}_A - \hat{\gamma}_B\right) + R_{T2} \qquad \text{(A.3)}$$

where $\bar{f}_r = \partial(\lambda_r)/\partial(\hat{\alpha}_r) = -\lambda_r^2 + \hat{\alpha}_r\lambda_r$, r=A, B, $\hat{\alpha}_r = \bar{Z}_r\hat{\gamma}_r$, and R_{T1}, R_{T2} are the residuals of approximation.

Using (A.3) and (A.1) the decomposition in (13) can be expressed as:

$$\begin{aligned}\overline{W}_A - \overline{W}_B = &\left[\left(\bar{X}_A - \bar{X}_B\right)\hat{\beta}_A + \hat{\theta}_A\bar{f}_A\left(\bar{Z}_A - \bar{Z}_B\right)\hat{\gamma}_A\right] \\ &+ \left[\bar{X}_B\left(\hat{\beta}_A - \hat{\beta}_B\right) + \hat{\theta}_A\bar{f}_B\bar{Z}_B\left(\hat{\gamma}_A - \hat{\gamma}_B\right)\right] + \left(\hat{\theta}_A - \hat{\theta}_B\right)\bar{\lambda}_B + R_M + R_{T1} + R_{T2}\end{aligned} \qquad \text{(A.4)}$$

The expression in (A.4) is then used to obtain the weights for the contribution of each characteristic and return as shown in (14). To obtain $P^{i}_{\Delta Z}$ and $P^{i}_{\Delta \gamma}$ as in (14) it should only be noted that $\hat{\theta}_A \bar{f}_A$ and $\hat{\theta}_A \bar{f}_B$ do not vary across the variables in Z.

<<<< INSERT TABLE A.1 ABOUT HERE >>>>

REFERENCES

ARRAZOLA M., DE HEVIA J., RISUEÑO M. and SANZ J. F. (2003) Returns to education in Spain: Some evidence on the endogeneity of schooling, *Education Economics* 11, 293–304.

ARRAZOLA M. and DE HEVIA J. (2008) Three measures of returns to education: An illustration for the case of Spain, *Economics of Education Review* 27, 266-75.

BECKER G.S. (1964) *Human capital: A theoretical and empirical analysis with special reference to education*, Chicago, The University of Chicago Press.

BLACKABY D. H. and MANNING N. (1990) The North-South divide: Questions of existence and stability?, *The Economic Journal* 100, 510–27.

BLACKABY D. H. and MURPHY P. D. (1991) Industry characteristics and inter-regional wage differences, *Scottish Journal of Political Economy* 38, 142–61.

BRONZINI R. and PISELLI P. (2009) Determinants of long-run regional productivity with geographical spillovers: The role of R&D, human capital and public infrastructure, *Regional Science and Urban Economics* 39, 187-99.

BAUM C.F., SCHAFFER M.E. and STILLMAN S. (2010) *ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression*. http://ideas.repec.org/c/boc/bocode/s425401.html

CAMERON A. C. and TRIVEDI P. K. (2005) *Microeconometrics: Methods and Applications*. Cambridge University Press, New York

CARD D. (1999) The causal effect of education on earnings, in: Ashenfelter O. & Card D. (Eds) *Handbook of Labor Economics*, Vol. 3, Amsterdam, Elsevier Science.

CARD D. (2001) Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems, *Econometrica* 69, 1127-60.

CICCONE A., CINGANO F. and CIPOLLONE, P (2004) The private and social return to schooling in Italy, *Giornale degli Economisti e Annali di Economia* 63, 413-44.

COMBES P., DURANTON G. and GOBILLON L. (2008) Spatial wage disparities: Sorting matters!, *Journal of Urban Economics* 63, 723–42.

DE LA FUENTE A., DOMÉNECH, R. and JIMENO, J.F. (2003) *Human capital as a factor of growth and employment at the regional level: The case of Spain*, Report for the European Commission, DG for Employment and Social Affairs.

DICKIE M and GERKING S. (1987) Interregional wage differentials: An equilibrium perspective, *Journal of Regional Science* 27, 577–85.

DI LIBERTO A. (2008) Education and Italian regional development, *Economics of Education Review*, 27, 94-107.

DUSTMANN C. and MEGHIR C. (2005) Wages, experience and seniority, *Review of Economic Studies* 72, 77-108.

FARBER S. and NEWMAN R. (1987) Accounting for South/Non-South real wage differentials and for changes in those differentials over time, *The Review of Economics and Statistics* 69, 215–23.

FARBER S. and NEWMAN R. (1989) Regional wage differentials and the spatial convergence of worker characteristic prices, *The Review of Economics and Statistics* 71, 224-31.

GARCÍA J., HENÁNDEZ P.J. and LOPEZ-NICOLÁS A. (2001) How wide is the gap? An investigation of gender wage differences using quantile regression, *Empirical Economics* 26, 149-67.

GARCÍA I. and MOLINA A. (2002) Inter-regional wage differentials in Spain, *Applied Economics Letters* 9, 209-15.

GARCÍA-PÉREZ J. I. and JIMENO J. F. (2007) Public sector wage gaps in Spanish regions, *The Manchester School* 75, 501–31.

GERKING S. and WEIRICK W. (1983) Compensating differences and interregional wage differentials", *The Review of Economics and Statistics*, 65, 483–487.

GREENE W. H. (2003) *Econometric Analysis* (5th ed.), Prentice-Hall, New Jersey

HARMON C. and WALKER I. (1995) Estimates of the Economic Return to Schooling for the United Kingdom, *American Economic Review*, 85, 1278-86.

HECKMAN J. (1979) Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-161.

HOFFMANN R. and KASSOUF A. (2005) Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure, *Applied Economics* 37: 1303-11.

ICHINO A. and WINTER-EBMER R. (1999) Lower and upper bounds of returns to schooling: an exercise in IV estimation with different instruments, *European Economic Review* 43, 889–901.

ICHINO A. and WINTER-EBMER R. (2004) The long-run educational cost of World War II, *Journal of Labor Economics* 22, 57–86.

KRUEGER A. B. and LINDAHL M. (2001) Education for growth: Why and for whom?, *Journal of Economic Literature* 39, 1101-36.

LÓPEZ-BAZO E. and MORENO R. (2008) Does human capital stimulate investment in physical capital?: Evidence from a cost system framework, *Economic Modelling*, 25, 1295-1305

MAIER G. and WEISS P. (1986) The importance of regional factors in the determination of earnings: the case of Austria, *International Regional Science Review* 10, 211–20.

MONTUENGA V., GARCÍA I. and FERNANDEZ M. (2003) Wage flexibility: Evidence from five EU countries based on the wage curve, *Economics Letters* 78, 169-74.

MOTELLÓN E. (2008) Un análisis de las diferencias regionales en el impacto de la contratación temporal en España, *Investigaciones Regionales* 12, 107-31

MURRAY M.P. (2006) Avoiding invalid instruments and coping with weak instruments, *Journal of Economic Perspectives* 20, 111-32.

NEUMAN S. and OAXACA R. (2004) Wage decompositions with selectivity-corrected wage equations: A methodological note, *Journal of Economic Inequality* 2, 3-10.

PEREIRA P. T. and SILVA P. (2004) Returns to education and wage equations, *Applied Economics* 36, 525–31.

REDDING, S. and SCHOTT P. K. (2003) Distance, skill deepening and development: will peripheral countries ever get rich? *Journal of Development Economics* 72, 515-41.

REILLY B. (1992) An analysis of local labour market wage differentials, *Regional Studies* 26, 257-64.

RODRÍGUEZ-POSE A and VILALTA-BUFÍ M (2005) Education, migration, and job satisfaction: The regional returns of human capital in the EU, *Journal of Economic Geography*, 5, 545-66.

SIMÓN H., RAMOS R. and SANROMÁ E. (2006) Collective bargaining and regional wage differences in Spain: an empirical analysis, *Applied Economics* 38, 1749-60.

STOCK J.H. and YOGO M. (2005) Testing for weak instruments in linear IV regression, in Andrews D.W.K. and Stock J.H. (Eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge, Cambridge University Press.

WILLIAMS N. (2009) Seniority, experience, and wages in the UK, *Labour Economics* 16, 272-83.

WOOLDRIDGE J. (2002) *Econometric analysis of cross section and panel data*, MIT Press, Cambridge MA.

YUN M. (2004) Decomposing differences in the first moment, *Economics Letters* 82, 275-80.

Table 1. Hourly wages and human capital in the Spanish regions.

	Obs.	Hourly gross wage (€)	Real Hourly gross wage (€)	Schooling (years)	Experience (years)	Tenure (years)				
						≤ 1	2-4	5-9	10-14	≥ 15
Andalusia	1336	6.60 (3.256)	6.91 (3.409)	9.65 (4.046)	17.65 (12.121)	39.07%	17.29%	7.86%	9.13%	26.65%
Aragon	576	8.00 (4.507)	8.29 (4.671)	11.16 (3.945)	19.45 (11.769)	21.70%	19.79%	9.90%	14.41%	34.20%
Asturias	396	7.06 (4.152)	7.00 (4.111)	10.22 (3.950)	17.58 (11.951)	29.55%	18.18%	8.59%	14.14%	29.55%
Balearic Isl.	379	6.80 (3.291)	6.50 (3.143)	10.07 (3.477)	19.28 (12.253)	41.95%	21.37%	7.12%	8.71%	20.84%
Basque Country	618	8.75 (4.362)	8.29 (4.134)	11.17 (3.761)	17.75 (11.785)	25.40%	20.06%	7.93%	12.46%	34.14%
Canary Isl.	848	6.49 (4.224)	6.65 (4.327)	9.43 (3.790)	17.22 (11.981)	37.38%	23.47%	8.25%	8.73%	22.17%
Cantabria	455	6.55 (3.963)	6.65 (4.023)	10.54 (3.667)	18.50 (12.416)	24.40%	23.96%	8.57%	9.45%	33.63%
Castile-La Mancha	613	6.53 (3.545)	7.06 (3.836)	9.38 (3.841)	17.89 (12.324)	36.05%	18.60%	10.60%	10.93%	23.82%
Castile-Leon	683	7.78 (4.659)	8.15 (4.879)	10.62 (3.917)	19.55 (11.726)	21.08%	21.23%	10.98%	11.13%	35.58%
Catalonia	1513	7.92 (4.490)	7.44 (4.216)	10.29 (3.870)	18.94 (12.153)	27.96%	21.08%	11.43%	12.43%	27.10%
Extremadura	482	6.10 (3.496)	7.05 (4.037)	9.81 (4.030)	17.16 (11.894)	31.74%	21.37%	8.30%	10.79%	27.80%
Galicia	795	6.42 (3.830)	6.51 (3.888)	10.03 (3.886)	17.45 (12.051)	28.30%	22.52%	12.20%	7.92%	29.06%
La Rioja	358	6.67 (2.932)	6.57 (2.886)	10.39 (3.525)	18.29 (11.349)	24.30%	21.23%	9.22%	13.13%	32.12%
Madrid	1174	8.49 (4.800)	8.51 (4.809)	11.89 (3.828)	18.10 (11.938)	24.28%	22.57%	11.33%	11.07%	30.75%
Murcia	558	6.23 (3.498)	6.37 (3.580)	9.44 (4.108)	17.92 (12.415)	29.75%	23.84%	9.68%	11.83%	24.91%
Navarre	496	7.91 (3.543)	7.36 (3.296)	10.43 (3.522)	19.39 (12.290)	25.81%	21.98%	7.86%	10.69%	33.67%
Valencia	886	6.37 (2.922)	6.46 (2.961)	9.85 (3.942)	18.27 (12.311)	30.47%	21.56%	9.03%	12.98%	25.96%
Spain	12166	7.19 (4.062)	7.24 (4.063)	10.27 (3.930)	18.23 (12.077)	29.67%	21.08%	9.62%	11.06%	28.58%

Note: Sample means and standard deviation, in parentheses, for the continuous variables. Share of each category for tenure.

Table 2. Description of the variables in the empirical wage model for Spain and for the regions with the highest and lowest wage levels.

	SPAIN		MADRID		MURCIA	
	Employees	**Non-Employees**	**Employees**	**Non-Employees**	**Employees**	**Non-Employees**
WORKER'S HUMAN CAPITAL						
Schooling (years)	10.274 (3.930)	7.512 (3.593)	11.887 (3.828)	8.321 (3.758)	9.443 (4.108)	7.361 (3.754)
Experience (years)	18.229 (12.077)	- -	18.098 (11.938)	- -	17.925 (12.415)	- -
Tenure						
≤ 1 year	29.67%	-	24.28%	-	29.75%	-
2-4 years	21.08%	-	22.57%	-	23.84%	-
5-9 years	9.62%	-	11.33%	-	9.68%	-
10-14 years	11.06%	-	11.07%	-	11.83%	-
≥ 15 years	28.58%	-	30.75%	-	24.91%	-
INDIVIDUAL AND FAMILY CHARACTERISTICS						
Age (years)	37.275 (10.691)	41.854 (12.605)	37.401 (10.468)	43.469 (11.892)	36.149 (11.279)	39.769 (12.578)
Household size (persons)	3.697 (1.276)	3.839 (1.368)	3.538 (1.205)	3.684 (1.177)	3.925 (1.324)	4.009 (1.466)
Other household income (€ per month)	1083.114 (974.891)	1407.772 (883.635)	1282.540 (1130.720)	1719.724 (1071.399)	1024.553 (1029.854)	1318.850 (830.680)
Nº children ≤ 15 years	0.753 (0.891)	0.896 (1.075)	0.691 (0.878)	0.889 (1.050)	0.928 (0.991)	1.215 (1.311)
Children 0-6 years	24.53%	28.13%	22.91%	27.98%	31.18%	34.21%
Gender						
Male	60.00%	18.15%	53.41%	11.24%	63.26%	17.41%
Female	40.00%	81.85%	46.59%	88.76%	36.74%	82.59%
Marital status						
Married	65.72%	76.88%	65.84%	83.40%	66.31%	77.81%
Other	34.28%	23.12%	34.16%	16.60%	33.69%	22.19%
Chronic disease	8.78%	21.97%	9.03%	21.15%	10.22%	22.34%

Note: Sample means and standard deviation, in parentheses, for the continuous variables. Share of each category for the discrete characteristics.

Table 3. Wage level within categories of worker human capital endowments.

	SPAIN	MADRID	MURCIA
Education			
Illiterate	5.18	5.60	4.49
Primary	5.73	6.33	5.11
Secondary	6.74	7.46	5.97
Tertiary	10.78	11.37	10.53
Experience			
≤ 1 year	4.70	5.05	3.94
2-9 years	5.92	6.80	5.41
9-19 years	7.47	8.87	6.75
19-29 years	8.12	9.08	6.94
≥ 30 years	8.06	10.12	7.11
Tenure			
≤ 1 year	5.37	6.04	4.56
2-4 years	6.32	7.50	5.70
5-9 years	7.28	8.69	5.79
10-14 years	8.38	9.78	7.66
≥ 15 years	9.38	10.67	8.80

Note: Sample mean of real wage per hour in € within each category.

Table 4. Estimated returns to education in the Spanish regions.

	IV HECKIT		OLS	IV	HECKIT	
	COND	UNCOND			COND	UNCOND
Andalusia	0.0819	0.2161	0.0409	0.0774	0.0619	0.1950
Aragon	0.0739	0.2050	0.0589	0.0743	0.0634	0.1979
Asturias	0.0124	0.1110	0.0373	0.0942	0.0347	0.1668
Balearic Isl.	0.0709	0.1804	0.0689	0.0471	0.0572	0.1670
Basque Country	0.0752	0.1813	0.0657	0.0861	0.0535	0.1560
Canary Isl.	0.0588	0.1500	0.0625	0.0758	0.0744	0.1636
Cantabria	0.0996	0.1711	0.0522	0.0755	0.0735	0.1463
Castile-La Mancha	0.0790	0.2112	0.0650	0.0950	0.0686	0.2148
Castile-Leon	0.0788	0.1864	0.0489	0.0633	0.0750	0.1831
Catalonia	0.0957	0.1765	0.0508	0.0682	0.0531	0.1251
Extremadura	0.1057	0.2916	0.0552	0.1007	0.0728	0.2593
Galicia	0.1032	0.2065	0.0580	0.1061	0.0814	0.1862
La Rioja	0.0780	0.1572	0.0457	0.1150	0.0447	0.1248
Madrid	0.0676	0.1846	0.0706	0.0607	0.0601	0.1619
Murcia	0.1057	0.2047	0.0445	0.0636	0.0671	0.1572
Navarre	0.0890	0.1517	0.0458	0.0814	0.0515	0.1063
Valencia	0.0873	0.2106	0.0545	0.0510	0.0502	0.1604
Spain	0.0785	0.1881	0.0541	0.0737	0.0617	0.1673

Notes: All the estimated returns in the table are statistically significant at 1%.
COND and UNCOND denote the conditional and unconditional returns respectively.

Table 5. Estimated returns to experience and tenure in the Spanish regions.

	Experience		Tenure							
			2-4 years		5-9 years		10-14 years		≥15 years	
Andalusia	0.0158	***	0.0684	**	0.0919	*	0.2116	***	0.1560	***
Aragon	0.0100	***	0.0903	**	0.1844	***	0.2162	***	0.3606	***
Asturias	0.0077		0.1012		0.1651		0.4472	***	0.4249	***
Balearic Isl.	0.0074	***	0.0291		0.0780		0.1300	**	0.1728	***
Basque Country	0.0104	***	0.1056	**	0.1831	***	0.3196	***	0.3462	***
Canary Isl.	0.0103	***	0.1031	***	0.1119	**	0.2848	***	0.3810	***
Cantabria	0.0078	*	-0.0464		0.0261		0.0557		0.1670	*
Castile-La Mancha	0.0101	***	-0.0305		0.1041	**	0.0762		0.2777	***
Castile-Leon	0.0135	***	-0.0280		0.1165	**	0.2694	***	0.2351	***
Catalonia	0.0172	***	0.0967	***	0.1727	***	0.2165	***	0.1900	***
Extremadura	0.0193	***	0.0473		0.0140		0.0910		0.0752	
Galicia	0.0181	***	0.0621	*	0.0069		-0.0039		0.1211	***
La Rioja	0.0116	***	0.1795	***	0.0699		0.2432	***	0.1910	***
Madrid	0.0103	***	0.1250	***	0.2238	***	0.3545	***	0.3783	***
Murcia	0.0126	***	0.0762		0.1151	*	0.2144	***	0.2543	***
Navarre	0.0127	***	0.0916	**	0.2353	***	0.2087	***	0.2746	***
Valencia	0.0129	***	0.0598		0.0650		0.1825	***	0.2491	***
Spain	0.0123	***	0.0746	***	0.1289	***	0.2182	***	0.2605	***

Notes: Returns computed from the IV Heckit estimates of the Mincerian wage specification. The omitted category for tenure is "≤ 1 year".

***, **, and * denote significant at 1%, 5%, and 10% respectively.

Table 6. Regional wage gap decomposition.

	Wage Gap	Global Decomposition			Contribution of Human Capital		Contribution of Schooling	
		Endowment	Return	Residual	Endowment	Return	Endowment	Return
Andalusia	0.0342	0.0648	-0.0015	-0.0290	0.0790	-0.1401	0.0542	-0.0712
Aragon	-0.1352	-0.1046	0.0960	-0.1266	-0.1107	-0.2416	-0.0712	-0.2819
Asturias	0.0415	-0.0037	0.8060	-0.7603	0.0055	-0.4598	0.0039	-0.4213
Balearic Isl.	0.0845	0.0408	0.1204	-0.0766	0.0329	-0.0454	0.0163	-0.0601
Basque Country	-0.1666	-0.0815	-0.0263	-0.0588	-0.0815	-0.0109	-0.0711	-0.0888
Canary Isl.	0.1146	0.1095	-0.0400	0.0451	0.1063	0.2159	0.0699	0.1829
Cantabria	0.0993	-0.0353	0.1990	-0.0644	-0.0337	-0.2258	-0.0209	-0.3090
Castile-La Mancha	0.0259	0.0877	0.2443	-0.3060	0.0903	-0.4908	0.0704	-0.4734
Castile-Leon	-0.1055	-0.0723	0.0307	-0.0639	-0.0679	-0.1449	-0.0279	-0.0967
Catalonia	-0.0456	-0.0017	-0.1250	0.0810	-0.0120	-0.1267	-0.0014	-0.0533
Extremadura	0.0307	0.0443	0.0374	-0.0510	0.0537	-0.5080	0.0366	-0.3677
Galicia	0.1201	0.0286	0.1613	-0.0697	0.0324	-0.4811	0.0210	-0.3477
La Rioja	0.0541	-0.0283	0.2544	-0.1721	-0.0279	-0.2657	-0.0095	-0.2751
Madrid	-0.1699	-0.1334	-0.0606	0.0241	-0.1471	0.2341	-0.1387	0.1972
Murcia	0.1366	0.0739	-0.2707	0.3333	0.0785	0.0885	0.0667	0.0478
Navarre	-0.0517	-0.0399	-0.1637	0.1519	-0.0376	0.1258	-0.0127	0.0555
Valencia	0.0896	0.0381	-0.1528	0.2043	0.0393	0.2292	0.0356	0.2385

Notes: Results for the regional wage gaps decomposition obtained from the IV HECKIT estimates. The wage gap is defined as the difference between the average (log) wage in the rest of the country and in the corresponding region. Global decomposition refers to the one in NEWMAN and OAXACA (2004) for models with sample selection. Contribution of Human Capital and Schooling refers to the part of the wage gap explained by differences in the endowment and return to human capital (schooling, experience, and tenure) and schooling respectively, from the detailed decomposition in eq. (14).

Table A.1. Complementary results of the IV Heckit estimation of the wage model.

	Observations		Exclusion Restr.[1]	Weak Identif. Test[2]	Underidentif. Test [3]	Sargan[4] (p-value)	IMR [5]
	Censored	Uncensored					
Andalusia	1903	1336	war lge chronic	13.14 (<15%)	38.596***	1.174 (0.278)	-0.040
Aragon	455	576	war lge age age^2 married chronic child0-15 child0-6 nhouse	23.92 (<5%)	160.145***	11.733 (0.110)	0.206*
Asturias	511	396	war lge age age^2 child0-15 child0-6	4.23 (>30%)	24.65***	1.256 (0.870)	1.220***
Balearic Isl.	302	379	war lge age age^2 chronic	13.85 (<10%)	60.291***	5.526 (0.137)	0.087
Basque Country	467	618	war lge age age^2	22.85 (<5%)	80.987***	4.440 (0.109)	0.044
Canary Isl.	765	848	war lge married chronic child0-15 child0-6 nhouse	8.03 (<20%)	53.577***	5.843 (0.322)	-0.158*
Cantabria	388	455	war lge age age^2 child0-15 child0-6	9.51 (<10%)	51.64***	7.578 (0.108)	0.056
Castile-La Mancha	743	613	war lge age age^2 child0-15 child0-6	26.09 (<5%)	126.988***	5.121 (0.275)	0.451***
Castile-Leon	670	683	war lge age age^2 married chronic child0-15 child0-6 nhouse inchouse	20.67 (<5%)	161.940***	10.389 (0.239)	0.036
Catalonia	943	1513	war lge age age^2	22.27 (<5%)	84.744***	1.935 (0.380)	-0.28
Extremadura	692	482	war lge age age^2	12.92 (<5%)	47.750***	0.445 (0.801)	0.004
Galicia	763	795	war lge age age^2 married chronic child0-15 child0-6 nhouse	18.32 (<5%)	138.41***	7.535 (0.375)	0.034
La Rioja	261	358	war lge age age^2 married chronic child0-15 child0-6 nhouse inchouse	11.39 (<10%)	89.838***	6.881 (0.549)	0.305***
Madrid	747	1174	war lge chronic nhouse	22.06 (<5%)	82.857***	1.208 (0.546)	-0.133
Murcia	649	558	war lge age age^2	5.41 (<30%)	21.262***	1.054 (0.590)	-0.619*
Navarre	269	496	war lge age age^2	8.74 (<10%)	33.416***	1.553 (0.460)	-0.479*
Valencia	795	886	war lge age age^2	10.04 (<10%)	38.922***	1.613 (0.446)	-0.497**
Spain	11323	12166	war lge chronic	221.91 (<10%)	623.52***	0.684 (0.408)	-0.076**

Notes: (1) Variables used as instruments in the wage equation. *War* and *lge* refer to the dummy variables for the Spanish Civil War and the General Education Act as defined in the text. *Chronic* denotes if the individual is affected by a chronic disease, *child0-15* the number of children younger than 15 years old, *child0-6* if there are children less than 6 years old, *nhouse* the number of members in the household, and *inchouse* the income of the household excluding the wage of the worker.
(2) Value of the Cragg-Donald Wald F statistic, and in ( ) the corresponding % of the maximal relative bias of the IV estimator over the OLS estimator under the null hypothesis of weak instruments. Due to the number of instruments used for Spain and Andalusia, the % corresponds to the desired maximal size of a 5% Wald test involving the coefficients of the two instrumented variables in the wage equation under the null hypothesis of weak instruments. In both cases, STOCK and YOGO (2005) provided the critical values.
(3) Value of the Anderson canonical correlation LM statistic, and significance from the corresponding $\chi^2(L)$ distribution with L the number of instruments.
(4) Sargan's overidentification test of all the instruments.
(5) Estimate of the coefficient of the Inverse Mills ratio (IMR) in the wage equation.
***, **, and * denote significant at 1%, 5%, and 10% respectively.

[1] An alternative is to estimate the effect of human capital on firms' productivity. However, the lack of firm-level data from a representative survey for each Spanish region prevented us from considering this approach. In any case, under well-known assumptions, the marginal productivity explanation of wage determination establishes the link between wages and productivity. The assessment of the return to human capital based on the estimation of a wage equation is standard in the labour market literature.

[2] The ECHP has frequently been used in wage studies for the Spanish labour market and for other EU Member States (MONTUENGA et al., 2003; RODRÍGUEZ-POSE and VILALTA-BUFÍ, 2005; GARCÍA-PÉREZ and JIMENO, 2007). Although the Earnings Structure Survey (a dataset also produced in the EU countries under the auspices of EUROSTAT) contains the most complete information on wages and workers, jobs and firms' characteristics, it does not provide information on the non-employed. This prevents us from controlling for sample selection and computing the indirect effects of human capital on wages through its effect on the probability of employment, which is one of the objectives of this paper.

[3] The regional representativeness of the sample for the entire panel of the ECHP is only guaranteed at the NUTS I level, which corresponds in Spain to an artificial grouping of regions based on geographical criteria alone.

[4] In any case, individuals working less than 15 hours a week were removed from the sample, given that in this case the ECHP does not provide information on some variables that are important for our analysis (e.g. tenure). It should be mentioned that the results are robust to the exclusion from the sample of those individuals working less than 30 hours a week.

[5] This information was kindly provided by the Catalan Institute for Statistics (IDESCAT), which estimates the parity power standards for the 17 Spanish regions from the aggregate Spanish figures used by the Statistical Office of the EU, EUROSTAT, to produce a data net of the cost of

living differences across the Member States. Note that, given the common currency for the spatial units under analysis, parity power standards only account for differences in the cost of living.

6 It can be argued that jobs' and firms' characteristics also differ across regions. And as far as wages vary within these characteristics, the composition effect should include them as well. However, here the focus is on individuals' characteristics, given our interest in the effect of human capital. In any case, a great deal of the wage variability associated with different jobs and firms is likely to be captured by differences in workers' human capital if there is a process of sorting across jobs and firms depending on the endowment of human capital.

7 Results for the 17 regions are not reported here to save space, though they are available upon request.

8 Note that, as is usual in this type of analysis, a simple specification of the Mincerian wage equation is used to obtain a better insight into the global effects of the human capital variables on wages (see PEREIRA and SILVA, 2004).

9 The full list of characteristics included in X and in Z is shown in Table 2. See section 4.2 for further details.

10 It must be noticed that this type of selection process does not distinguish between non-participants and unemployed individuals. It is possible to design a two-step sequential selection procedure for the decision of participation and then for employment for those participating (ARRAZOLA and DE HEVIA, 2008 considered this type of selection process in their study of the returns to schooling in Spain). However, we decided not to use such a process due to the low number of observations for some of the categories in some regions and, above all, given our interest in the derivation of a detailed decomposition of the regional wage gaps. In any case, results reported in ARRAZOLA and DE HEVIA (2008) suggest that the estimate of the returns

does not vary markedly when the selection process distinguishes between non-participants and unemployed.

[11] See GREENE (2003) and CAMERON and TRIVEDI (2005) for the derivation of the expressions and the discussion of these marginal effects. HOFFMANN and KASSOUF (2005) and ARRAZOLA and DE HEVIA (2008) used these expressions to compute different types of returns to education.

[12] As usual in the specification of wage equations, a quadratic form is used for experience ($\beta_{EXP} \cdot EXP_{ir} + \beta_{EXP2} \cdot EXP^2_{ir}$). As a result, the return to experience (conditional and unconditional) is $\beta_{EXP} + 2 \cdot \beta_{EXP2} \cdot EXP_{ir}$. In the case of tenure, its return will be measured by the estimation of the coefficients of each of its categories.

[13] Despite the arguments suggesting the overestimation of the returns to schooling based on the OLS estimator, the conclusion from the results in the literature based on IV estimates is that the causal effect of education is as big or bigger than the OLS estimated return.

[14] GARCÍA et al. (2001) also used a similar instrument for their analysis of the gender wage gap in Spain.

[15] Results of the Sargan test were clearly against the exogeneity of the inverse Mills ratio. This is an important difference when comparing our results for the entire country with the ones reported in ARRAZOLA and DE HEVIA (2008), as they did not take into account that the inverse Mills ratio is a function of human capital, and thus considered it as an exogenous regressor. We thank an anonymous referee for raising that point.

[16] We thank an anonymous referee for raising this concern.

[17] A two-step procedure was implemented in STATA 11. In the first step, the inverse Mills ratio was estimated by using the command HECKMAN. In a second step, we used the IVREG2 routine (BAUM et al., 2010) to estimate the wage equation by TSLS, considering schooling and the inverse Mills ratio as endogenous regressors.

[18] The full set of estimated coefficients for the wage and the selection equations, for each of the 17 regions and for Spain, are available from the authors upon request.

[19] These results are available from the authors upon request.

[20] Notice that in what follows it is assumed that the no-discrimination wage structure is that in region A.

[21] Notice that $P^i_{\Delta X}$ and $P^i_{\Delta \beta}$ are the weight in the standard linear decomposition.

[22] It is impossible to summarize the results for the decomposition of the wage gap for all pairs of regions (17*16*0.5=136) in this type of publication. An alternative to the one in our study is to consider the gap with regard to a benchmark region (for instance, the one with the highest average wage), although this is subject to the criticism of the selection of the benchmark and slightly complicates the comparison of results across regions. In any case, the qualitative conclusion on the important contribution of regional differences in the return to human capital (in particular to schooling) is also obtained when a benchmark region is used to compute the gaps. These results are available from the authors upon request.