Educational Spillovers in Portugal*

Sandra Sousa[†1], Miguel Portela[1,2,3] and Carla Sá[1,2,4]

[1]School of Economics and Management, University of Minho
[2]Economic Policies Research Unit (NIPE)
[3]Institute for the Study of Labor (IZA)
[4]Centre of Research in Higher Education Policies (CIPES)

March 8, 2015

Abstract

Using a Portuguese panel dataset and employing instrumental variables in an augmented Mincer specification, in this paper is analysed the external effects of human capital on individual wages, where the aggregated human capital is measured by three different ways: regional average education, share of qualified workers in regional workforce and a regional skill index. The results propose the existence of positive and significant external effects of human capital. In addition, it is considered two qualification groups of workers as imperfect substitutes and it is found that the effect is larger for the most qualified workers.

Furthermore, the effect on individual wages slightly decreases when the aggregated human capital of firm is included as an additional control and it seems that human capital externalities predominate within firms compared to local human capital in a county.

Finally, assessing the relationship between geographical (regional and inter-regional) human capital externalities across the mainland Portuguese counties and individual wages. Results indicate that beyond the individual education and the human capital level of the county where individuals are working, the human capital level of the neighbouring counties also matters.

JEL classification: C23, C26, D62, J24.

Keywords: Panel data models; Instrumental variables estimation; Externalities; Human capital.

*Support provided by the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia) under the grant SFRH/BD/85985/2012 is gratefully acknowledged.

†School of Economics and Management, University of Minho, 4710-057 Braga, Portugal
E-mail:sipsousa@hotmail.com

1 Introduction

Human capital externalities have been deemed to be a key determinant of growth and development (Lucas, 1988), i.e. human capital investments have positive externalities that foster economic growth and, as such, these externalities determine the optimal subsides to education. However, while much attention has been devoted to the empirical investigation on the role of human capital in the economic growth process, as well as on the private returns of human capital, only a few works have attempted to estimate the external effects of human capital at the individual level. The assess of the existence and magnitude of human capital externalities, it has been the focus of some authors (Rauch, 1993; Acemoglu and Angrist, 2001; Moretti, 2004a,b,c; Ciccone and Peri, 2006; Dalmazzo and de Blasio, 2007; Heuermann, 2011), but there is no consensus about the findings. Rauch (1993) proposes the existence of aggregate human capital externalities within the United States metropolitan areas and Moretti (2004a) suggests the existence of human capital externalities for United States cities, i.e. the relationship between individual earnings and the concentration of human capital can be explained partly by the productivity gains that workers may obtain by interaction with more skilled workers, in locations with a high concentration of human capital. However, more recent papers have obtained opposite results, providing no evidence of local human capital externalities for the United States cities (Ciccone and Peri, 2006), and for the American states (Acemoglu and Angrist, 2001).

In this essay, it is intended to answer questions related to the external returns to human capital, such as: Do human capital externalities matter for individual earnings in Portugal? If there are, to what extent they are more or less relevant than individual educational attainment? The goal is to assess if there are human capital externalities, at different aggregation levels, and what their magnitudes. Since, the most existing literature uses data from the United States and literature on this issue is almost unknown in Portugal, this essay contributes to increasing and understanding how the interactions among Portuguese workers and regions can be determinants of differences in individual wages.

In this sense, different from most of the existing empirical literature, external returns to human capital are analysed in three different perspectives: at the regional level (Rauch, 1993; Moretti, 2004a); at the firm and regional level (Bauer and Vorell, 2010) and, finally, at the regional and inter-regional level, considering the possibility of spatial externalities between regions (Rodríguez-Pose and Tselios, 2012).

So, in this analyse it is estimated augmented Mincerian wage equations using a Portuguese matched employer-employee database, wherein the aggregated human capital is measured by three different ways: regional average education, regional share of high qualified workers in region and an adjusted multi-dimensional skill index (Portela, 2001) of the region.

Taking into account the unobserved time-invariant individual characteristics and the endogeneity of aggregate human capital, it can be suggested the existence of positive external effects of human capital. Following Moretti (2004a), the presence of a polytechnic institute in the county is used as instrument to regional human capital and it is proposed a new instrument, the share of people who could read in the county according to the 1900 census.

The remainder of this essay is organized as follows. The next section provides a literature review on external effects of education and section 3 provides a simple theoretical framework. Section 4 describes the empirical identification strategy and data used for estimation. Section 5 reports the estimation results for three groups of workers: all workers, most qualified workers and less qualified workers, and Section 6 concludes.

2 Literature Review

Along several decades, various authors have focused in research about the effects of human capital on the individual earnings, being practically consensual that, the greater the educational level the higher the private gains of the workers. However, according with empirical literature, there are some findings indicating that social returns to education exceed private returns and one of the reasons mentioned is the presence of externalities. So, the central idea of theoretical explanations is that the cumulative knowledge by agents in a local environment may benefit the other agents, i.e. the intuition is that the average stock of human capital in a region could have an effect on individual wages of the workers in this region (Rauch, 1993; Acemoglu and Angrist, 2001; Moretti, 2004a,b; Ciccone and Peri, 2006; Lange and Topel, 2006; Dalmazzo and de Blasio, 2007; Heuermann, 2011). Thus, the strength of human capital externalities is defined as the difference between the marginal social return and the private return to an additional unit of human capital (Garcia-Fontes and Hidalgo, 2009).

In empirical literature, there are three main strategies to evaluate the external effects of human capital: to study the effects of the aggregate human capital on individual wages in a city or region, which is the strategy used in this essay; to assess the effects of the aggregate human capital on productivity indicators using data at firm level or at regional level; to evaluate the effect of the education on social outcomes, such as, crime rates (Moretti, 2004b; Lange and Topel, 2006).

Knowledge of the existence of human capital spillovers in economic theory is ancient. Marshall (1890) was the first to recognize that the interactions among workers create learning opportunities which may yield increases in productivity. The most recent literature has been built upon this Marshall's vision and have been proposed some models where the human capital externalities are the fundamental engine to economic growth (Aziariadis and Drazen, 1990; Lucas, 1988; Black and Henderson, 1999; Lange and Topel, 2006). Lucas (1988), in a prominent article, suggests that the externalities of human capital in the shape of learning spillovers, may be large enough to explain the differences in the long-run income across poor and rich countries. Furthermore, the issue on human capital externalities is also extremely relevant, because determine which magnitude of human capital accumulation that must be financed (Heckman and Klenow, 1998; Heckman, 2000; Isacsson, 2005; Garcia-Fontes and Hidalgo, 2009).

Despite significant political implications, the empirical evidence on existence and magnitude of external returns to education are very scarce, when compared with the extensive literature about private returns to education (Card, 1999, 2001). Just recently, some authors attempted to estimate the magnitude of the education spillovers by comparing the wages of similar people who work in cities or states with different average education level (Rauch, 1993; Acemoglu and Angrist, 2001; Moretti, 2004a,b; Ciccone and Peri, 2006; Heuermann, 2011).

Mincerian wage equation augmented with an aggregated human capital term at region level is the approach employed by the most authors (for instance, Rauch, 1993; Rudd, 2000; Acemoglu and Angrist, 2001; Moretti, 2004b; Ciccone and Peri, 2006). This methodology estimates the strength of the human capital externalities by observing the effect of regional human capital level in individual wages (Garcia-Fontes and Hidalgo, 2009). The basic idea is that the human capital externalities should appear in the individual wages, since all relevant individual characteristics are controlled, such as education, gender, experience, among others.

In a pioneer work, Rauch (1993) uses a cross-sectional data for United States Standard Metropolitan Statistical Areas for 1980 and employs the average level of formal education, as well, work experience as proxies of average level of human capital. The author finds significant and positive effects of average level of schooling on individual wages, i.e. an increase of an year of average schooling in the Metropolitan Area raises the workers' wages between 3% and 5%. Concerning the average work experience, the author found a weak and insignificant coefficient.

Several authors have augmented and improved the Rauch's work (1993), namely, Acemoglu and Angrist (2001) which use pooled data from United States Census, to the period 1960 - 1980, and find similar results to those obtained by Rauch (1993), when they use OLS estimation. Outside the United States, some researches suggest the existence of positive and significant human capital externalities: Sweden (Isacsson, 2005), Italy (Dalmazzo and de Blasio, 2007), Spain (Garcia-Fontes and Hidalgo, 2009), China (Liu, 2007), Russia (Muravyev, 2008), Netherlands (Canton, 2009) and Germany (Bauer and Vorell, 2010; Heuermann, 2011), where the OLS estimates of an extra year in the average level of schooling is associated with an increase, between 0.6% and 6.7%, on average wages .

However, in literature there are two main criticism of the Rauch's work, since, on the one hand, the author doesn't argue the endogeneity of human capital variable, and on the other hand, doesn't distinguish between external effect and complementary effect among skilled and unskilled workers (Moretti, 2004b; Ciccone and Peri, 2006). Consequently, the dimension of the coefficient of human capital spillovers may be biased due to the presence of unobservable factors that are correlated with the average stock of human capital and wages among the areas (Moretti, 2004a; Dalmazzo and de Blasio, 2007). The key feature in the estimation of the human capital externalities, at regional level, is that changes in levels of aggregate human capital are endogenous due the migration among regions, since regions with high productivity and higher wages may attract more skilled workers. Moreover, regions with high income might have amenities which are especially attractive for the highly skilled workers. Thus, to reduce this bias, it is desirable to implement an instrumental variables approach (Moretti, 2004a; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009).

In the instrumental variables methodology, a good instrument for the average schooling must affect the schooling of the majority of workers in a location, but without being correlated with the local wages (Duranton, 2006). Thus, the literature proposes a variety of instruments: changes in the child labour laws and compulsory education laws at the state level (Acemoglu and Angrist, 2001), lagged demographic structure of the city/region (Moretti, 2004a; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009), existence of colleges in cities (Moretti, 2004a,b), regional share of workers with a college degree 20 years before the observed period (Bauer and Vorell, 2010) and the number of high schools of the region, as well as the number of students attending them (Heuermann, 2011).

Results are not consensual in empirical studies, where the instrumental variables approach is used. Acemoglu and Angrist (2001), using pooled data of the Census from 1960 to 1980, emphasize insignificant effects to the external returns to education, but when added data for the year 1990, the effect becomes significant, indicating the rise in importance of human capital. Unlike, Moretti (2004a) and Iranzo and Peri (2009), using the regional share of high skilled workers as instrument, in the United States, find substantial evidences to the human capital externalities. Moretti (2004a,b) develops a theoretical model to capture the external effects of the education on wages, employing individual data to United States and taking into account the endogeneity of the choice of location, instrumenting the share of high skilled workers by the existence of land-grant colleges in a region. Author finds a wage premium of 0.4% to the college educated workers if the share of the high educated workers increase by one percentage point and to low skilled workers the estimate is considerably higher varying between 1.6% and 1.9%. In this work, in certain specifications and regions, it is found a spillover effect of over than 9%.

In literature, the estimates of educational externalities found, vary with time period under study, level spatial aggregation, country and specification. In addition, the impact of the average education in region and the regional share of skilled on wages is completely extinguished, when the human capital stock of the firms is included as a control variable. In this case, evidence is not found for the human capital spillovers at the regional level (Canton, 2009).

Moreover, a key assumption in the Mincerian regressions is that workers with different human capital level are perfect substitutes in production (Moretti, 2004a,b; Ciccone and Peri, 2006; Canton, 2009; Garcia-Fontes and Hidalgo, 2009). In other words, it assumes that changes in the relative supply of human capital doesn't affect the relative wages of the different human capital groups, keeping the total factors productivity constant. However, empirical evidence indicates that different levels of human capital are imperfect substitutes (Angrist, 1995; Katz and Murphy, 1992; Ciccone and Peri, 2006; Hidalgo, 2009). Hence, Thus, all effects caused by changes in the relative supply of human capital, on workers with a given level of human capital, have to come through total factor productivity and should be interpreted as externalities(Moretti, 2004b; Ciccone and Peri, 2006).

If different levels of human capital are imperfect substitutes, the Mincerian approach yields a positive effect of aggregate human capital on individual wages (Ciccone and Peri, 2006). The imperfect substitutability implies that an increase in aggregate supply of skilled workers tends to increase the unskilled workers wages and diminish the wages of high level education workers. Moretti (2004a) investigates the role of imperfect substitution by estimating of Mincerian regressions for separated groups to education. Ciccone and Peri (2006) have introduced a constant-composition approach to estimate the gross spillovers to correct the bias, and it allows the identification of human capital externalities even when the workers with different educational levels are imperfect substitutes. Thus, this approach estimates the strength of the human capital externalities as the marginal effect of the aggregate levels of human capital on average wages holding constant the composition of the workforce (Garcia-Fontes and Hidalgo, 2009), even if the skilled and unskilled workers are imperfect substitutes. The authors suggest that the external effects of education on wages are identified through the variation of the average educational level across different points of time in a given geographical area, keeping constant the weights of skills composition in this geographical area. Ciccone and Peri (2006) use pooled data at the regional level, for the period between 1970 and 1990, and they find no evidence of external returns to the average years of education in US cities and states. On the other hand, using the same approach, Iranzo and Peri (2009) find no external returns for the years of education at secondary level, but a substantial return to the college level. Literature reports that Mincerian approach yields larger externalities than the constant-composition approach, i.e.

the first approximation provides an upper bound for the human capital externalities while the second provides a lower limit.

In Portugal, there are some studies documenting the existence of substantial wage gap between different regions (for instance, Cardoso, 1991; Teulings and Vieira, 2004; Vieira et al., 2005, 2006) and it is pointed out that the wages are higher in the region of Lisbon and Tagus Valley comparatively to the other regions of the country.

Nevertheless, empirical studies about education externalities at the regional level are unknown, and only the working paper of Silva (2003) is found. Silva (2003), employing an empirical approach with data for workers who were displaced between different counties over the period from 1989 to 1999, finds positive evidence of educational spillovers at the county level, which range between 1.7% and 8.4%. However, controlling for individual fixed-effects, the results suggest that an increase of one year in the average schooling in a county implies a small positive impact on individual wages (around 1.7%), and when the firms controls are included, this result flocks to zero. The author also proposes that, after 5 years of displaced, the wages of a college graduate workers increase by more than 3% comparing to a workers with basic schooling, if there is an increase of one year in the average education of their county.

Also in context of Portuguese literature about this issue, but at the firm level, Martins and Jin (2010) find evidence of high returns, that range between 14% and 23%, implying that the social returns to education are higher than the private returns, and this situation is due to a transfer of skills from the workers with more formal education to the workers with less formal education. They also find evidence of significant wage spillovers to less-educated workers, since their pay increases by 2% to 3% per extra year of education of workers in their firm, but not to their more educated colleagues.

3 Econometric Framework

3.1 Estimation Strategy

The starting point of the analysis is to estimate, by OLS, the effect of the aggregate human capital at the regional level on individual hourly wage by exploiting a Mincerian wage equation augmented with a local human capital term, given by:

$$\log W_{irt} = \beta X_{irt} + \eta \bar{S_{rt}} + \delta Z_{rt} + d_{sa} + d_t + \mu_{irt} \quad (1)$$

with W_{irt} denoting the hourly wage of individual i in county r at time t; X_{irt} is a vector of observable characteristics of individual i in county r at time t, including gender, education, experience and tenure, and their corresponding squares, as well as, a set of dummies to occupation and a firm size control; $\bar{S_{rt}}$ represents regional human capital in the county r in time t; Z_{rt} is a matrix of county characteristics, and d_{sa} and d_t control to the sector of activity and time fixed-effects, to capture the unobserved time-invariant factors of sector of activity and time that may influence the wages; and μ_{irt} is the error term to the unobservable characteristics and it is assumed to have the the basic assumptions of the classic regression model. The variable of interest is $\bar{S_{rt}}$ and the estimated coefficient of this variable provides evidence of the existence of regional spillover effects of education, thus, the main goal is to estimate η, the impact of local human capital on individual wages.

Since different ways of measuring regional human capital can be characterized by different spillovers, three forms to measure the regional human capital are distinguished. So, two of these measures have been generally used in previous studies, namely average years of education in the region (as in Rauch, 1993; Acemoglu and Angrist, 2001; Silva, 2003; Dalmazzo and de Blasio, 2007; Liu, 2007) and share of highly skilled workers in the region workforce (as in Moretti, 2004a; Garcia-Fontes and Hidalgo, 2009; Bauer and Vorell, 2010). In this analyse, the skilled workers are those with 12 or more years of schooling, i.e. secondary and college educational levels.

To provide a contribution to the existing literature, a measure to the aggregate human capital based in the skill index proposed by Portela (2001) is used. This skill index take in account different dimensions of the productivity of the worker, namely schooling, labour experience market and unobservable ability, which in this analysis it is adjusted at the county level. So, the skill index of a given county with n workers is given by:

$$SI_{county} = mschool \left(0.5 + \frac{e^{\frac{\overline{E_c} - mschool}{sschool}}}{1 + e^{\frac{\overline{E_c} - mschool}{sschool}}} \right) \times \left(0.5 + \frac{e^{\frac{\overline{Exper_c} - mexper}{sexper}}}{1 + e^{\frac{\overline{Exper_c} - mexper}{sexper}}} \right)$$

where $\overline{E_c} = \dfrac{\sum_{i=1}^{n} school_i}{n}$ is the average schooling of the county and $school_i$ is the schooling level of worker i, evaluated as years of schooling; *mschool* represents the average of school level in the population and *sschool* denote the standard deviation of schooling; $\overline{Exper_c} = \dfrac{\sum_{i=1}^{n} exper_i}{n}$ is the average experience of the county and $exper_i$ is the experience of the worker i, *mexper* and *sexper* represent the average and the standard deviation of experience, respectively. In addition, the same skill index is computed at the firm level.

Additional control variables at the county level that may affect the concentration of human capital and wages are included. Thus, following the empirical literature a proxy to unemployment rate in the county is included, as a measure of local unemployment, since the correlation of education with earnings might be affected by the distribution of unemployment across counties. If the most skilled individuals are less likely to be unemployed, then the aggregate human capital variable may capture the effect of the unemployment rate. This variable also captures the local conditions of the labour market in different periods. In addition, a control variable for the average firm-size at the county level is included, which is an index of local competitiveness (Dalmazzo and de Blasio, 2007). The compensatory effect of amenities on wages is controlled in the model including the attractiveness index of the county, which is correlated with the welfare of the individuals, since the greater the positive amenities of a county the higher its attractiveness[1]. Regions with better amenities tend to attract more skilled workers, thus, with higher potential, and may be those who are better able to learn, which make them more productive (Rauch, 1993; Moretti, 2004a,b; Combes et al., 2011). Furthermore, the impact of local human capital on wages could also reflect agglomeration effects (Dalmazzo and de Blasio, 2007), since the larger the dimension of the local population potentially presents benefits of agglomeration associated with better market linkages and, therefore, might make more productive workers (Heuermann, 2011). So, to control the local effects of the local characteristics in that act as agglomeration economies, the log population of the county is included. Finally, to control the regional price level, the ratio of the mean price of evaluation of accommodation per square meter between each county and county of Lisbon is included, since this variable measures the influence of changes in the cost of living at the county level in the period

[1] Heuermann (2011) use the number of hotel beds in a county as a proxy for a region's amenity endowment, since a large number of hotel beds indicates that it is attractive for people to travel into that region, be it out of leisure or business motives.

considered.

However, there are some methodological issues that must be addressed. So, the OLS estimation of Mincerian equation to the identification of human capital externalities (1) may lead to biased estimates of the coefficients of interest variables. On the one hand, there may be unobserved factors that are correlated with individual wages and education, i.e. one source of bias is associated with the omission of unobservable characteristics of workers that influence their productivity and, therefore, the wage. By other hand, due the correlation between wages and human capital in a county, for example, workers born in regions with the best schools and universities, and thus, integrated into an environment favourable to accumulation of human capital are more productive (Combes et al., 2011) or migration across counties implies that regional schooling levels are endogenous (Moretti, 2004a,b). Therefore, if controls for unobservable characteristics of workers are not included in the wage equation, possibly part of the estimated wage differential associated with the concentration of human capital can be explained by the fact that these regions have a higher number of skilled workers, i.e. regions that demand a high-skilled workforce may also offer higher returns to unobserved abilities, leading to upward biased estimates of η.

The panel structure of data allows to control for individual time-invariant unobserved factors that may be correlated with education, eliminating part of the bias which can influence the coefficients associated with human capital variables. There are several benefits from using panel data models, namely, they give more informative data, variability, degrees of freedom and efficiency, and less collinearity among the variables; it allows to control for individual heterogeneity, i.e. panel data suggests that individuals are heterogeneous and, so, it is able to control for time-invariant variables, and since the fixed-effects method only estimates within effects then may not suffer from heterogeneity bias (Baltagi, 2008).

Therefore, it is intended to estimate by fixed-effects method the following:

$$\log W_{irt} = \alpha_i + \beta X_{irt} + \eta \bar{S_{rt}} + d_{sa} + d_t + \epsilon_{irt} \qquad (2)$$

in which $\mu_{irt} = \alpha_i + \epsilon_{irt}$. α_i represents the random individual-specific effect that characterizes each worker and it is constant through time, and it is independent of ϵ_{irt}, which is varying over time and individuals, and it is assumed to be uncorrelated with all included variables. Thus, it is also assumed that ϵ_{irt} is a sequence of i.i.d. random variables with mean zero and variance σ_ϵ^2, $N(0, \sigma_\epsilon^2)$, and that α_i is i.i.d. over the panels with mean zero and variance σ_α^2, $N(0, \sigma_\alpha^2)$.

An important characteristic of this model is that some variables are correlated with the individual-specific effect, for instance, it is assumed that unobserved ability of the individual is included in α_i, which may be correlated with schooling and wage of the individual, assuming that unobserved ability of each individual doesn't vary over time. In a panel data framework, this situation leads to fixed-effects estimation, in which the only assumption that is required to assume is that the correlation between any variable and the error is through the individual-specific fixed-effect.

Otherwise, if it is assumed that all variables included in the model are uncorrelated with α_i and ϵ_{irt}, the estimates are both consistent and efficient using Feasible Generalized Least Squares, often called by random-effects estimation. However, take into account that some variables are endogenous, individual educational level and regional human capital, it seems more appropriate,in this analysis, the application of the fixed-effects estimator, because this correlation becomes the random-effects estimator inconsistent.

Note that, in specification 2, control variables for the characteristics of the county are no included, because in the sample only a negligible number of workers moved across counties (represent only about 5.7% of the sample), thus, individual fixed-effects estimates allow to control for unobserved time-invariant characteristics of individuals, as well as to the counties where individuals are working, because the control variables at county level do not vary across individuals.

Furthermore, in this model the set of time variables is dropped, since including time dummies for all years, it is not possible to estimate the effect of any variable whose shift over time is constant, as the experience variable, *Experience*, which is an individual specific time trend. The *Experience* can be expressed by the following:

$$Exper_{i,t} = Exper_{i,0} + (t-1), \quad t = 1, ..., T$$

which is is a linear combination, where each individual works every year, such that the *Experience* increases one unit each year, for each individual of the sample, and as the experience increases in the same amount for all individuals, the effect of increasing one year of experience can not be distinguished from aggregated temporal effects (Wooldridge, 2010; Greene, 2012).

Another alternative reported in the empirical literature to deal with the problem of endogeneity is the instrumental variable (IV) approach, i.e. these problems can be jointly tackled when there is an instrument for local human capital and another for individual education. The advantage of the IV method is that a valid instrument isolates the effect of exogenous changes in human capital levels on wages (Moretti, 2004a,b). As previously mentioned, the OLS estimates might still be biased due to endogenous location choices of individuals and unobservable characteristics of the county, thus, a key issue when estimating human capital externalities at the regional level is that changes in aggregate human capital levels are endogenous as counties with higher productivity and wages may attract more high skilled workers (Moretti, 2004a,b; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009). Another factor that may work in the same direction is that high income regions may have amenities that are especially attractive for high skilled workers. In this case, the coefficient of local human capital should be upward biased. In order to correct for local supply and demand shocks of skilled workers, the IV2SLS approach is applied. What is needed is an instrument that is correlated with human capital in a region and uncorrelated with unobserved factors that affect wages directly (Moretti, 2004a), the instrument must be taken into account for the observed variation in the local human capital, but not correlated with the residual of the earning equation, i.e. should be orthogonal to current individual wages.

In this essay, two alternatives are considered. Following Moretti (2004a), the first instrument proposed is the presence of a Polytechnic Institute in a county. The second instrument proposed is the share of people who can read in the county based on the data of Census 1900, which is something new in empirical literature.

According to literature, the presence of a Polytechnic Institute in a county tends to increase the education of this county, since the costs to attend the higher education are smaller (there are no costs for housing, for example) and people tend to fix itself to work where they live and study (Moretti, 2004a). In order to eliminate doubts about the exogeneity of instrument, it is used as instrument the presence of a Polytechnic Institute in a county that were established by the so-called educational reform of *Veiga Simão*, in the 70s of the twentieth century.

In 1971, Minister of Education *Veiga Simão* presented a school system project proposing deep structural changes in the Portuguese education system. This political project was based on the expansion and diversification of the education, in which it is created a new modality of higher education - the polytechnic higher education. This new modality of higher education arises from the need of training of intermediate graduated technicians for industry, agriculture and other services, and look for readjust access to higher education to more socially democratic levels. By other words, the goal is to create equal opportunities towards the education regardless of social origins and the region where it is inserted (Grácio, 1998). After the revolution of April 25 in 1974, the Law 513 - T79, of December 26, publishes a network of Polytechnic Institutes to deploy which included 27 high schools in 17 counties district headquarters, covering all the national territory. The decision process for the creation of this network takes into account the development needs of the regions where the Polytechnic Institutes would be deployed, that would be achieved through the training of technicians with practical preparation, and specific scientific and economic areas of each region. This instrument may be valid, since workers in a county with a Polytechnic Institute are not systematically different from other workers with the same education in a county where there is no a Polytechnic Institute. And, in addition, decisions for the installation process of the polytechnic education system were political and took place for nearly three decades ago. So, the presence of a Polytechnic Institute is likely to be uncorrelated with unobservable factors that affect wages in period under study. For all these factors, the geographical location of Polytechnic Institutes seems to be the close random can be considered (Moretti, 2004a).

Concerning the second instrument proposed, the central idea is based on the intergenerational influence of taste for study, which follows the same reasoning of the empirical studies that use characteristics of the family background as instrument to individual education (Card, 1995, 1999). The share of the people who can read in a county in 1900 is likely uncorrelated with unobservable factors that affect wages during 2002-2009, since this instrumental variable is based on data with more than a century ago. On the other hand, this variable affect the education of the county, because in counties where there are more literate people in 1900 and where there are probably a greater number of families able to provide education for its members through financial support, makes the county more educated, and the current education of the county, in principle, was affected by the influence the valuation of education that passes from generation to generation. Thus, the share of the people who can read in the county in 1900 may be used as a reliable instrument for human capital in the county.

Moreover, individual education is also treated as endogenous variable and the change in compulsory school attainment in 1981 that established the compulsory education of 9 years is used as instrument. Changes in compulsory schooling used as instrument is common in empirical economics, although, a bias may still arise, unless all components of unobserved ability are captured by the compulsory schooling (see Angrist and Krueger, 1991; Acemoglu and Angrist, 2001).

As previously, the specification to estimate by IV method includes the individual and firms characteristics, control dummies to sector of activity and time effects, using the different ways to measure human capital in county, average years of education, share of highly skilled workers and skill index. Since the panel structure is available, in order to complete this empirical analysis the instrumental variables fixed-effects estimation (IV-FE) is applied, to circumvent the potential endogeneity problem of human capital of the county. Note that, the control variables for the county are not used due to its collinearity with the instrument for the aggregate human capital, since the instrumental variables are included in order to control for the impact of unobservable time-region specific factors (Heuermann, 2011).

In addition, when the analysis of the existence of externalities at the firm level and at the interregional level is made, the OLS and fixed-effects estimation are used.

Another difficulty with identifying the causal impact of local human capital concentration on wages is that workers with different levels of education are imperfect substitutes in production (Moretti, 2004a,b; Ciccone and Peri, 2006; Heuermann, 2011). The intuition is that, with imperfect substitution, an increase in the number of skilled workers implies an increase of the wages of unskilled workers, that more than offsets the decrease in the wage of skilled workers (Moretti, 2004a,b; Ciccone and Peri, 2006; Garcia-Fontes and Hidalgo, 2009). Whenever, workers with different levels of education are imperfect substitutes in production, the parameter η will comprise effects that can be obtained by "composition effects", due to a larger share of skilled workers which increases the productivity of both types of workers, and by pure spillovers, due to human capital externalities. As a consequence, literature indicates that the estimated value of η must still be positive, even if spillovers are zero (Dalmazzo and de Blasio, 2007). According to the empirical literature, this role of imperfect substitution is studied by estimating regressions for separate education groups.

So, in this essay, the two groups of workers are formed according to the literature (for instance, Moretti, 2004a; Dalmazzo and de Blasio, 2007; Bauer and Vorell, 2010; Heuermann, 2011), which usually considers a group of less qualified workers, those do not have higher education, and highly qualified, otherwise. However, the data used in this essay, only provides two levels for higher education, 15 years of schooling (*bacharelato*) and 16 or more (graduation, master and doctoral degree), in which those who have a *bacharelato* represent only a small proportion of workers with higher education (18.8%), so, there is a low variability of the observations. Given that, since the estimation process requires variability of the data, then workers with the 12th grade are also included to the group of the most qualified workers. Thus, the groups formed are, the less qualified, constituted by those with 12 years (or less) of schooling and the second group, the most qualified, which are those with 12 years of schooling or more.

Finally, take in account the advantages mentioned above and the uncertainty associated with the choice of instruments to the regional human capital variables, panel data with instrumental variables is applied, namely the IV-type model proposed by Hausman and Taylor (1981). The selection of this procedure is motivated by several reasons. The fixed and random effects estimators are not suitable, since the aim is to obtain consistent and efficient estimates of both individual and external returns to human capital. As previously mentioned, due the existence of endogenous variables in the model, the fixed-effects estimator is the more suitable, because it produces consistent estimates, however, it is probably not fully efficient. In addition, an important advantage of the Hausman-Taylor (HT) approach is that one does not have to use external instruments, since the individual means of the strictly exogenous regressors are used as instruments for the time-invariant regressors that are correlated with the individual effects. Thus, Hausman-Taylor estimation is only possible if there are at least as many exogenous time-varying variables as endogenous time-invariant variables(Baltagi et al., 2003; Verbeek, 2008; Wooldridge, 2010). If is the case, the choice of the strictly exogenous time-varying variables is a testable hypothesis, then a Hausman test can be computed to compare the Hausman-Taylor and fixed-effects estimators, by other words, this test compares the consistent but inefficient fixed-effects estimates of parameters with the Hausman-Taylor estimates, which are consistent and efficient under the null hypothesis (Baltagi et al., 2003).

3.2 Data description

Data used in this essay are taken from four Portuguese distinct sources. Individual data on wages, age, gender, education, experience and further controls, as well as, firm data, such as sector of activity and size, are provided by *Quadros de Pessoal* (Personal Records), which is an unbalanced panel over the period 2002–2009 including firms and workers in mainland Portugal.

This panel dataset was augment by regional variables of the 278 mainland Portuguese counties, over the period 2002–2009, that may affect the concentration of human capital, as well as, the wages. The unemployment rates at county level are not available. Thus, a proxy for unemployment rates, using information of working population (Census 2001 and 2011) and the number of unemployed registered at Portuguese Employment Centres, is computed, and these data are provided by *Instituto Nacional de Estatística* and *Instituto de Emprego e Formação Profissional - Ministério do Trabalho e Solidariedade Social.* The information at county level used to population and general development index - index of attractiveness - are provided by Marktest Consulting - Sales Index 2013[2]. Data of 1900 Census and information of the mean price of evaluation of accommodation per square meter by county is provided by *Instituto Nacional de Estatística.*

The details of the variables [3] used in this essay are reported in the following table:

[2]Salex Index 2013 is a database annually updated, organized at the county level, which Marktest Consulting develops since 1992. This database results from continuous contact with official entities, as well as, with several private entities whose collaboration allows to offer relevant county level indicators, namely related to macroeconomics, demographics, business structure, economic activity and indices of regional development created exclusively by Marktest Consulting.

[3]See Table 11 in Appendix A for a definition of the variables.

Table 1: Descriptive statistics

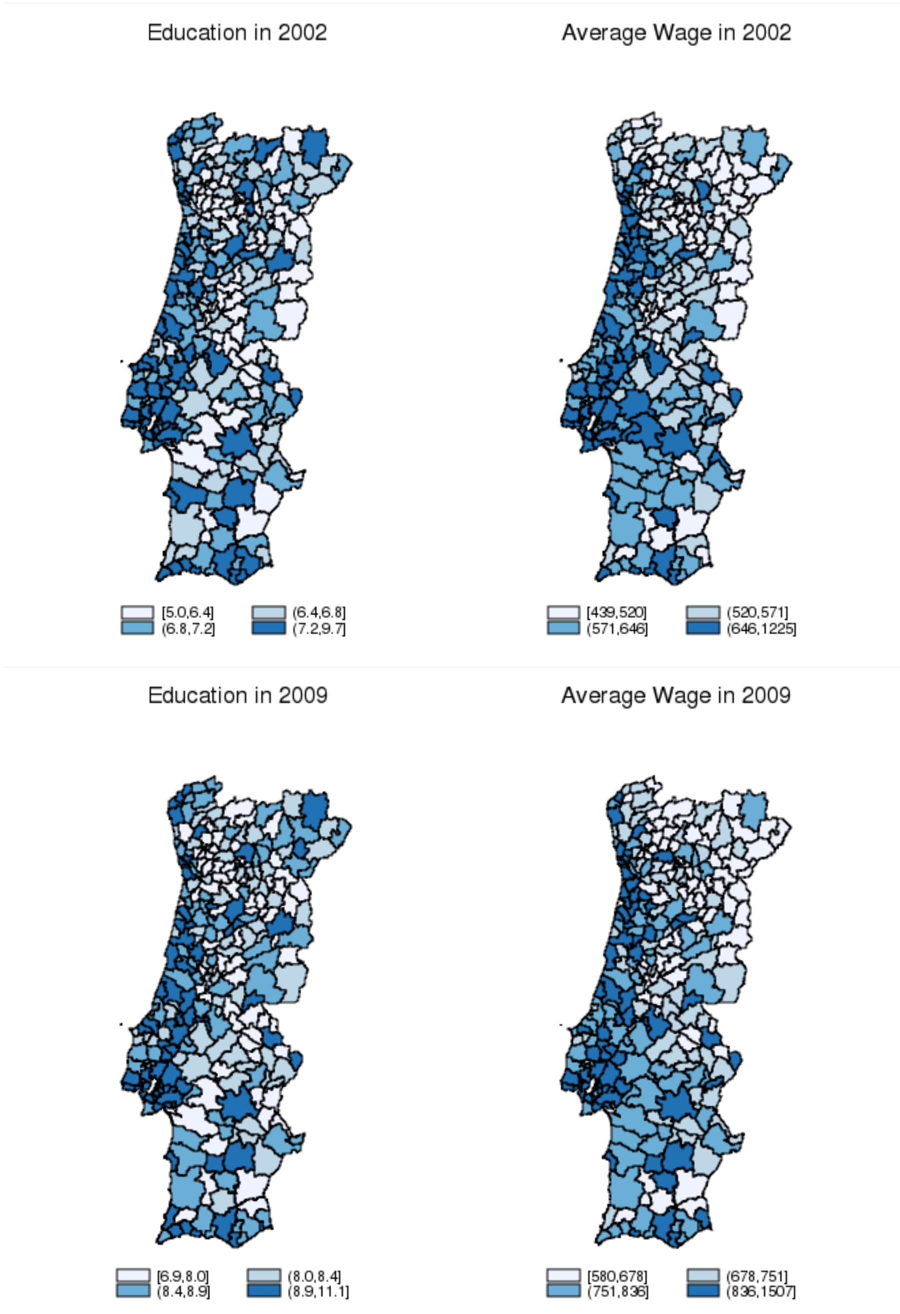
Variable	Obs.	Mean	Std. Dev.	Min	Max
Individual level variables					
Year	14 618 884	2006	2.270	2002	2009
Gender	14 618 884			0	1
Education	14 618 884	8.679	3.970	0	16
Age	14 618 884	38.25	10.81	15	65
Occupation	14 529 497			1	9
Lhwreal	14 618 884	1.557	0.568	0.233	7.237
Tenure	14 618 884	7.944	8.352	0.0833	53.83
Experience	14 615 739	21.09	11.30	0	49
Firm level variables					
Activity sector	14 257 794			2	21
Logsize	14 618 884	4.063	2.285	0	9.902
County level variables					
Log population	14 618 884	10.93	1.175	6.714	12.61
Average schooling by county	14 618 884	8.679	1.328	5.033	11.103
Share most-qualified workers by county	14 618 884	34.195	14.386	4	60.320
County Skill Index	14 618 884	8.8098	2.0096	3.4475	12.1536
Attractiveness Index (log form)	14 594 589	6.5234	1.7409	1.5810	8.8326
Housing price (ratio)	14 618 884	63.8886	21.6502	32.49	100
Unemployment rate	14 618 884	4.235	1.557	0.274	21.98
Log aversize	14 618 884	2.5349	0.4232	1.3743	3.1965
Number of workers	3 580 368	3 580 368	3 580 368	3 580 368	3 580 368

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting - Sales Index 2013, *Instituto Nacional de Estatística* and, *Instituto de Emprego e Formação Profissional - Ministério do Trabalho e Solidariedade Social*

Figure 1 shows both wage and educational inequalities, between Portuguese counties, at the beginning and end of the period under study, 2002 and 2009, respectively.

It turns out that, there are marked educational and wage asymmetries between the coastline and the interior of Portugal, which remains throughout the years. It is found that, Portuguese counties workforce education positioned in the lower quartile of the distribution of education are characterized by a lower schooling (average schooling up to 6.4 in 2002 and to 8.0 in 2009) and are located in the interior of the country. Contrary, those positioned in the highest quartile, in which the average schooling is higher than 7.2 (in 2002) and 8.9 (in 2009) are located in the coastline. With such differences in education levels at county level, it does not come as a surprise that average wages diverge across Portuguese regions, since practically every counties where the average education is higher coincide with the counties where wages are also higher.

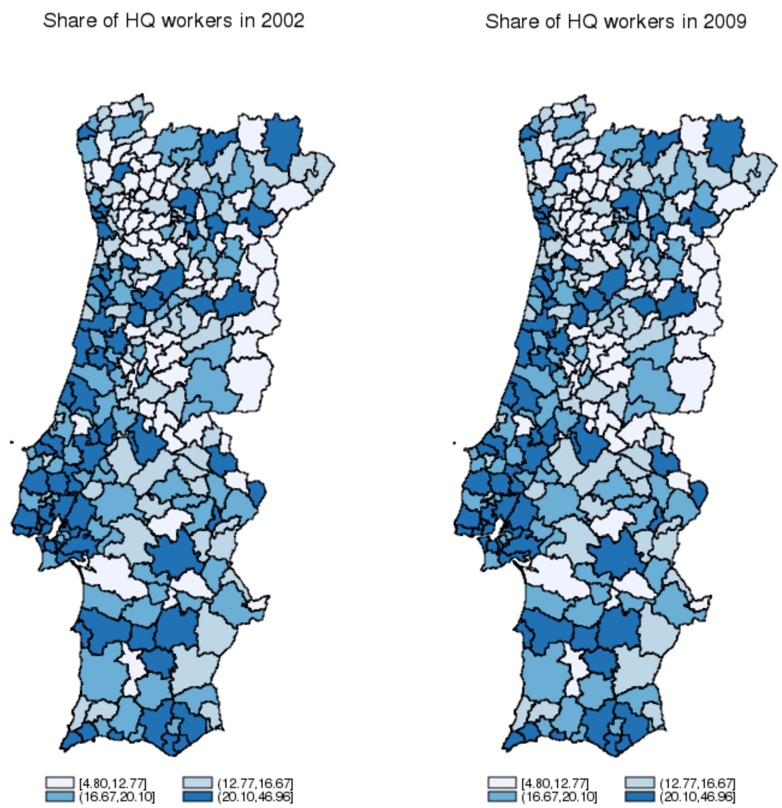
Figures 2 and 3 present a perspective of the distribution of human capital by county measured by the share of high qualified workers within a county's workforce, who have 12 years of schooling or more, and by the skill index which take in account different dimensions



Source: Created by author based on *Quadros de Pessoal*

Figure 1: Average Schooling and Average wages, by county, in 2002 and 2009

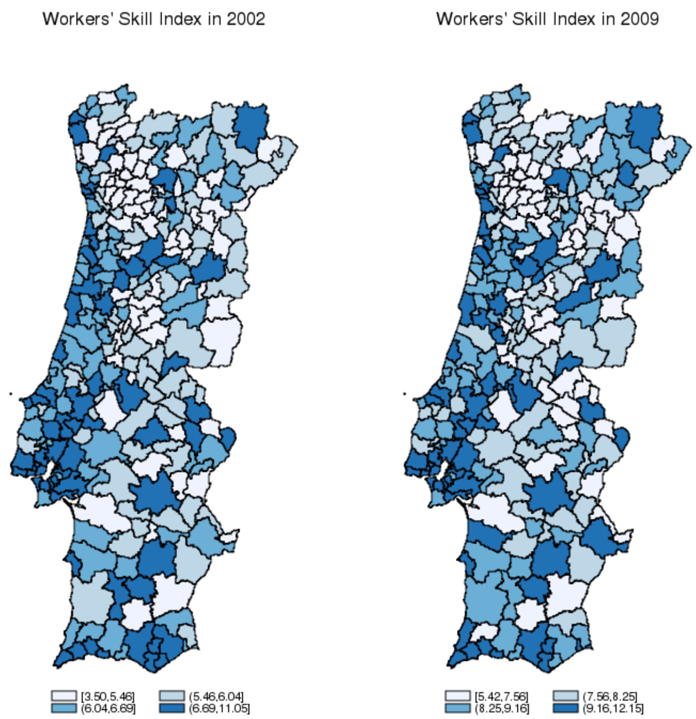
of productivity of the worker, respectively.



Source: Created by author based on *Quadros de Pessoal*

Figure 2: Educational attainment by county, as percentage of most qualified workers in 2002 and 2009

Relative to Figure 2, the share of the most qualified workers increases with the number of workers in a county (the linear correlation coefficient of Pearson is 0.81 in 2002, 0.76 in 2009 and 0.74 in period 2002–2009), and this share ranges from 4.7% to 47.2%, in 2002, and from 12.3% and 60.3%, in 2009. It is noted that, counties with a low values of share of the most qualified workers (below the 1st percentile) are mostly located in interior of Portugal, such as, *Arronches, Cinfães, Freixo de Espada à Cinta, Gavião, Paços de Ferreira, Tarouca* and *Vizela* in 2002, and *Felgueiras, Lousada, Marvão, Pampilhosa da Serra, Penamacor* and *Vizela* in 2009. In contrast the counties with highest value of share



Source: Created by author based on *Quadros de Pessoal*, 2002-2009

Figure 3: County Skill Index, in 2002 and 2009

of the most qualified workers are Lisbon followed by *Oeiras*, in both 2002 and 2009.

With respect to Figure 3, skill index by county ranges between 3.50 and 11.05, in 2002, and between 5.42 and 12.15, in 2009. The counties which have a value for the index below the 1st percentile are *Baião*, *Barrancos*, *Cinfaẽs*, *Felgueiras*, *Manteigas*, *Tarouca* and *Vizela*, in both 2002 and 2009. So, the counties with lower skill index are the same, excluding *Manteigas* and *Tarouca*, when the human capital is measured by the average education and by the share of the most qualified workers. As before, the counties with the highest skill index are *Oeiras*, following by Lisbon. Also there is a positive correlation between this skill index and the number of workers in a county (the linear correlation coefficient of Pearson is 0.80 in 2002, 0.74 in 2009 and 0.75 in period 2002–2009).

Note that, regardless the measure of the aggregate human capital used, in general, it appears that there is a higher concentration of human capital on coastline of Portugal than on the interior, and this asymmetry remains over years.

4 Empirical results

4.1 Human capital externalities at county level

The starting point of the analysis is essentially an extension of the Mincer approach in empirical labour economics. Thus, it is estimated wage equation which includes a set of individual characteristics, namely, individual education measured in years of schooling, a dummy for gender, labour market experience and its squared, tenure and its squared, and a set of dummies for occupation variable, as well as, variables that characterize the firms, such as, a set of dummies for activity sector and (log) size of firm. It also includes a variable of human capital by county and set of dummies that control for time effects.

As previous mentioned, the regional human capital is measured by three ways, average years of education, share of high qualified workers and the skill index in the county, since different ways of measure of regional human capital can be characterized by different spillovers. In order to control heteroscedasticity and possible serial correlation between firms, in all estimations robust standard errors clustered at the firm level are applied.

Results from estimating equation (1) by pooled OLS and fixed-effects are reported in Table 2, using the whole sample. Columns 1-6 report OLS estimates of the specification using the different ways to measure regional human capital: average years of education in the county (columns 1 and 3), share of most-qualified workers (columns 2 and 4) and skill index of the county (columns 3 and 6). Columns 7 to 9 report the fixed-effects estimates, using different ways of measuring local human capital by the previous order. In this case, control variables for the characteristics of the county are no included, because in the sample only a negligible number of workers moved from county, thus, individual fixed-effects estimates allows to control for unobserved time-invariant characteristics of individuals, as well as counties where individuals work, because the control variables at county level do not vary across individuals and are stable over time. As explained in section 3.1, specifications of columns 7-9 do not include the set of time dummies.

Table 2: Estimation results for pooled OLS and Fixed-effects estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pooled OLS						Fixed-effects		
VARIABLES	**Dependent variable: Log of real hourly wage**								
Educ	0.0484***	0.0485***	0.0491***	0.0483***	0.0483***	0.0488***	0.0132***	0.0132***	0.0137***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0001)	(0.0001)	(0.0001)
Average education	0.0457***			0.0296***			0.0122***		
	(0.0020)			(0.0036)			(0.0001)		
Share of HQ workers		0.0042***			0.0031***			0.0013***	
		(0.0002)			(0.0003)			(0.0000)	
Skill index (county)			0.0248***			0.0159***			0.0068***
			(0.0011)			(0.0019)			(0.0001)
Female	-0.1639***	-0.1637***	-0.1636***	-0.1611***	-0.1612***	-0.1609***			
	(0.0033)	(0.0033)	(0.0033)	(0.0036)	(0.0036)	(0.0036)			
Exper	0.0206***	0.0206***	0.0205***	0.0207***	0.0207***	0.0207***	0.0298***	0.0298***	0.0306***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)
Exper squared	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0004***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0163***	0.0163***	0.0163***	0.0159***	0.0159***	0.0159***	0.0060***	0.0061***	0.0062***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0000)	(0.0000)	(0.0000)
Tenure squared	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	0.0474***	0.0473***	0.0475***	0.0459***	0.0459***	0.0459***	0.0261***	0.0263***	0.0265***
	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0001)	(0.0001)	(0.0001)
Log aversize				0.0213***	0.0214***	0.0220***			
				(0.0047)	(0.0048)	(0.0048)			
Log population				0.0035	0.0045**	0.0041*			
				(0.0023)	(0.0023)	(0.0023)			
Log attractiveness index				-0.0316***	-0.0373***	-0.0321***			
				(0.0052)	(0.0054)	(0.0053)			
Housing price (ratio)				0.0017***	0.0016***	0.0017***			
				(0.0003)	(0.0003)	(0.0002)			
Unemployment rate				-0.0043***	-0.0042***	-0.0044***			
				(0.0009)	(0.0009)	(0.0009)			
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.9394***	1.1843***	1.0990***	1.1287***	1.3158***	1.2295***			
	(0.0320)	(0.0278)	(0.0286)	(0.0358)	(0.0421)	(0.0384)			
Observations	14,166,402	14,166,402	14,166,016	14,142,939	14,142,939	14,142,553	14,166,402	14,166,402	14,166,016
R-squared	0.640	0.640	0.640	0.641	0.641	0.641			
RMSE	0.341	0.341	0.341	0.341	0.341	0.341			
σ_α							0.4517	0.4519	0.4538
σ_ϵ							0.1492	0.1491	0.1491
ρ							0.9017	0.9018	0.9025
F test							18.84	18.80	18.77
Prob > F							0.0000	0.0000	0.0000

Source: Computations of the author
Notes: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1;
For FE model F test is the F test that all $\alpha_i = 0$

Results of Table 2 show that, for all estimated specifications, the coefficients obtained for both individual and firm characteristics control variables, present the expected signs with statistical significance at 1% level. Thus, the positive coefficient of individual returns to education is in accordance with the human capital theory, indicating that each individual year of schooling increases wages by about 5%. Tenure and experience display the usual non-linear impact on wages, so, labour market experience increases wages up to about 35 years of experience. Results also show that, as expected, we are facing a labour market where gender discrimination is still the norm, because women's wages are about 16% lower than men's wages with similar levels of educational attainment. Furthermore, compared with industry of "Mining", the results show that "Financial intermediation" is the only activity sector with wage premium above the base sector (this result is not reported in table), and wages are increasing in the size of the firm.

Respect to the variable of interest, local human capital, it is found strong evidence to the existence of human capital externalities, since the three coefficients of regional human capital variables are positive and highly significant in all specifications. Controlling both individual and firm characteristics, a unit increase in average education of the county is associated with a 4.6% increase in wages, by other side, an increase in the share of highly qualified workers of county by one percentage point increases wages by 0.42%. These results are in line with empirical literature (for instance Rauch, 1993; Silva, 2003; Moretti, 2004a; Isacsson, 2005; Canton, 2009; Heuermann, 2011). The county skill index also show that an additional year of calibrated education in the county leads to a wage premium of about 2.5%.

The results described so far, do not control adequately for unobserved county-specific characteristics that, as mentioned above, may be correlated with aggregated education of the county. Thus, these estimates are potentially biased or inconsistent. To solve, or at least mitigate, the omitted variable bias associated with the estimates in columns 1–3, explanatory variables measuring county-specific characteristics, that could affect both the concentration of human capital and wages, are included. The results are reported in columns 4, 5 and 6 of the Table 2.

The introduction of these additional controls reduces slightly the estimated external returns to education, but they remain highly significant, which confirms the existence of human capital spillovers. This is a less than 1.5 percentage point reduction from the estimate reported in columns 1 to 3 without such local controls. This result suggests that there is a positive bias in estimations without inclusion of county controls, indicating unobservable local idiosyncrasies, that affect the demand curve of the labour market are relevant to the OLS estimation of local human capital spillovers. Most of the county-specific variables are shown to be significant determinants of individual earnings. As expected, county unemployment significantly reduces wages (as in Dalmazzo and de Blasio, 2007; Heuermann, 2011) and the coefficient of the proxy variable to local amenities also presents a significant and negative sign, being in agreement with the theory of compensatory wage differentials (Roback, 1982), which emphasizes that more amenable locations affect positively the individual welfare and negatively the production costs (eg, polluted air), thus, the wages should be lower. By other words, it seems that workers are willing to accept a lower wage in exchange for more welfare. Conversely, earnings tend to be higher in counties where the index of local competition is higher, i.e. in counties where the average firm-size is higher. The coefficient of the county house price variable is also significant and positive, because the demand of workers for more attractive regions is reflected in an increase in the rental price of land (Roback, 1982), and consequently, in the housing price. Finally, regarding the size of the county, it can be observed that only in columns 5 and 6 the coefficients are statistically significant at level 5% and 10%, respectively. These results support the hypothesis that higher counties pay higher wages, i.e. the size of local population makes workers more productive (Dalmazzo and de Blasio, 2007).

Note that, the coefficients obtained for the control variables to the individual characteristics and the firms, remain unchanged compared to the previous regressions.

In addition, to the these last specifications an interaction term between individual education and local education measure (using the three measures of local human capital used previously) is added. It is verified that, if individual does not have school then do not benefits from educational externalities, since the coefficient of local human capital variable is not statistically significant in any of the specifications[4]. Therefore, to benefit from externalities of education the individual must have at least 8 months of schooling.

Moreover, considering that in the case of a high dispersion of education within county,

[4]See the results reported in Table 12, columns 1-3, in Appendix A.1.

i.e. higher deviations from the mean implies that the average education of the county may not be a good measure for local education. As such, new specifications are tested in which the measures of local human capital are replaced: the average education of the county is replaced by the coefficient of variation[5] (in logarithmic form) and the share of the most qualified workers is replaced by the Hirschman-Herfindahl Index[6] of education[7], which addresses the idea that human capital externalities are rooted in the local concentration of workers of the same educational background (Heuermann, 2011). From the results, it appears that greater heterogeneity (dispersion) of education within the county has a positive effect on wages, however, with a significance level of 10%, i.e. the wage premium is higher in counties with higher dispersion of education. On the other hand, following Heuermann (2011), the Hirschman-Herfindahl index measures the concentration of workers at most 12 years of schooling, which constitutes about 65% of the Portuguese workforce, thus, the higher the concentration of these workers in the local labour force, smaller the fraction of most qualified workers therein. So, the negative sign of this index might stem from a lower share of most-qualified workers in the workforce of the county, and this impact is close to zero. Furthermore, given that an increase of this index (higher concentration) has a negative effect on wages means that the two types of workers are complementary and not substitutes.

Columns 7, 8 and 9 of Table 2 report individual fixed-effects estimates. Note that, because gender is a time-invariant variable, the variable is dropped when calculating the first difference. Moreover, the positive and significant private return to education relates only to workers who increased their schooling over the period under study.

The results show a positive and significant effect of local human capital on individual wages. The F test indicates that there are significant unobserved individual effects, implying that pooled OLS would be inappropriate. In addition, the Hausman test's null hypothesis, that the random-effects estimator is consistent and efficient, is rejected (p-value Hausman test is 0.0000), i.e. the unobservable individual effects are correlated with the explanatory variables which makes the fixed effects estimation the most suitable model.

[5]coefficient of variation is given by the ratio between *standard deviation of education in the county* and *average education in the county.*

[6]Hirschman-Herfindahl Index of education is determined by the sum between the squared of share of non-qualified workers and squared of share of high-qualified workers

[7]See the results reported in in Table 12, columns 4 and 5, in Appendix A.1.

As previously mentioned, the external effects of human capital arise from two sources. Unskilled workers benefit of positive external returns that arise from the complementarity between workers, as well from the effect of the true spillovers. However, external returns to education for skilled workers can be either positive or negative, depending if the magnitude of the positive effect of the true spillovers is higher or lower than the negative effect of the increase in their aggregate supply. So, the estimates of external returns reported so far, may just be a reflection of the imperfect substitutability, since they are obtained with the full sample, i.e. with all individuals with different educational qualifications together. Thus, these results only present the average external effect of human capital across different educational groups.

In order to identify and measure suitably the impact of the human capital spillovers on wages, it is estimated the external returns to education for two skill groups, as in empirical literature (for instance Moretti, 2004a; Dalmazzo and de Blasio, 2007; Canton, 2009; Heuermann, 2011). As defined above, the first group, the less-qualified, is constituted by those with 12 years of schooling or less and, the second group, the most-qualified, are those with 12 years of schooling or more (high school and college).

Table 3 reports the pooled OLS and fixed-effects estimates for two groups of individuals and considering the three measures of local human capital previously used.

Table 3: Less qualified versus most qualified workers

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	**Dependent variable: Log of real hourly wage**					
	Measure of Human Capital used: Average education of the county					
	Less-qualified			Most-qualified		
	Pooled OLS		FE	Pooled OLS		FE
Education	0.0333***	0.0330***	0.0079***	0.0997***	0.0997***	0.0517***
	(0.0005)	(0.0005)	(0.0001)	(0.0020)	(0.0020)	(0.0002)
Average education	0.0435***	0.0301***	0.0085***	0.0618***	0.0442***	0.0136***
	(0.0019)	(0.0036)	(0.0001)	(0.0027)	(0.0056)	(0.0002)
Female	-0.1725***	-0.1699***		-0.1433***	-0.1425***	
	(0.0035)	(0.0039)		(0.0038)	(0.0041)	
Experience	0.0177***	0.0178***	0.0263***	0.0296***	0.0297***	0.0406***
	(0.0002)	(0.0002)	(0.0001)	(0.0005)	(0.0005)	(0.0001)
Experience squared	-0.0002***	-0.0002***	-0.0002***	-0.0004***	-0.0004***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0152***	0.0151***	0.0058***	0.0229***	0.0221***	0.0066***
	(0.0005)	(0.0005)	(0.0000)	(0.0008)	(0.0008)	(0.0001)
Tenure squared	-0.0002***	-0.0002***	-0.0001***	-0.0004***	-0.0004***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	0.0451***	0.0442***	0.0265***	0.0505***	0.0489***	0.0261***
	(0.0026)	(0.0026)	(0.0001)	(0.0035)	(0.0035)	(0.0001)

Continued on next page

Table 3 – continued from previous page

Log aversize	No	0.0113**	No	No	0.0308***	No
		(0.0047)			(0.0063)	
Log population	No	0.0053**	No	No	0.0000	No
		(0.0022)			(0.0038)	
Log attract. index	No	-0.0401***	No	No	-0.0222***	No
		(0.0052)			(0.0078)	
Housing price (ratio)	No	0.0019***	No	No	0.0013***	No
		(0.0003)			(0.0004)	
Unemployment rate	No	-0.0042***	No	No	-0.0025*	No
		(0.0009)			(0.0014)	
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	No	Yes	Yes	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.8948***	1.1200***		0.0068***	0.1623***	
	(0.0312)	(0.0351)		(0.0521)	(0.0587)	
Observations	12,381,330	12,358,931	12,381,330	4,870,546	4,866,578	4,870,546
R-squared	0.557	0.559		0.635	0.636	
RMSE	0.314	0.314		0.394	0.394	
σ_α			0.3736			0.4832
σ_ϵ			0.1424			0.1670
ρ			0.8731			0.8932
F test			17.33			19.83
Prob > F			0.0000			0.0000
	Measure of Human Capital used: Share of High Qualified workers in a county					
	Less-qualified			Most-qualified		
	Pooled OLS		FE	Pooled OLS		FE
Education	0.0334***	0.0331***	0.0079***	0.0997***	0.0997***	0.0518***
	(0.0005)	(0.0005)	(0.0001)	(0.0020)	(0.0020)	(0.0002)
Share of HQ workers	0.0039***	0.0030***	0.0009***	0.0055***	0.0043***	0.0014***
	(0.0002)	(0.0003)	(0.0000)	(0.0002)	(0.0005)	(0.0000)
Female	-0.1723***	-0.1698***		-0.1432***	-0.1427***	
	(0.0035)	(0.0039)		(0.0038)	(0.0041)	
Experience	0.0177***	0.0177***	0.0265***	0.0296***	0.0297***	0.0408***
	(0.0002)	(0.0002)	(0.0001)	(0.0005)	(0.0005)	(0.0001)
Experience squared	-0.0002***	-0.0002***	-0.0003***	-0.0004***	-0.0004***	-0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0152***	0.0151***	0.0058***	0.0229***	0.0221***	0.0066***
	(0.0005)	(0.0005)	(0.0000)	(0.0008)	(0.0008)	(0.0001)
Tenure squared	-0.0002***	-0.0002***	-0.0004***	-0.0004***	-0.0004***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	0.0451***	0.0442***	0.0262***	0.0505***	0.0489***	0.0258***
	(0.0026)	(0.0026)	(0.0001)	(0.0035)	(0.0035)	(0.0001)
Log aversize	No	0.0115**	No	No	0.0312***	No
		(0.0047)			(0.0063)	
Log population	No	0.0063***	No	No	0.0015	No
		(0.0022)			(0.0038)	
Log attract. index	No	-0.0448***	No	No	-0.0293***	No
		(0.0053)			(0.0082)	
Housing price (ratio)	No	0.0018***	No	No	0.0013***	No
		(0.0002)			(0.0004)	
Unemployment rate	No	-0.0042***	No	No	-0.0024*	No
		(0.0009)			(0.0014)	
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	No	Yes	Yes	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.1279***	1.3031***		0.3416***	0.4327***	
	(0.0269)	(0.0420)		(0.0498)	(0.0634)	
Observations	12,381,330	12,358,931	12,381,330	4,870,546	4,866,578	4,870,546
R-squared	0.557	0.560		0.635	0.636	

Continued on next page

Table 3 – continued from previous page

RMSE	0.314	0.314		0.394	0.394	
σ_α			0.3735			0.4826
σ_ϵ			0.1424			0.1670
ρ			0.8731			0.8931
F test			17.30			19.78
Prob > F			0.0000			0.0000

	Measure of Human Capital used: Skill index of county					
	Less-qualified			Most-qualified		
	Pooled OLS		FE	Pooled OLS		FE
Education	0.0340***	0.0335***	0.0082***	0.1005***	0.1003***	0.0526***
	(0.0005)	(0.0005)	(0.0001)	(0.0020)	(0.0020)	(0.0002)
Skill index (county)	0.0235***	0.0161***	0.0048***	0.0339***	0.0246***	0.0090***
	(0.0011)	(0.0018)	(0.0001)	(0.0015)	(0.0033)	(0.0001)
Female	-0.1722***	-0.1696***		-0.1432***	-0.1422***	
	(0.0035)	(0.0039)		(0.0038)	(0.0041)	
Experience	0.0176***	0.0177***	0.0271***	0.0295***	0.0297***	0.0414***
	(0.0002)	(0.0002)	(0.0001)	(0.0005)	(0.0005)	(0.0001)
Experience squared	-0.0002***	-0.0002***	-0.0003***	-0.0004***	-0.0004***	-0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0152***	0.0151***	0.0058***	0.0229***	0.0220***	0.0066***
	(0.0005)	(0.0005)	(0.0000)	(0.0008)	(0.0008)	(0.0001)
Tenure squared	-0.0002***	-0.0002***	-0.0001***	-0.0004***	-0.0004***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	0.0452***	0.0442***	0.0264***	0.0507***	0.0489***	0.0258***
	(0.0026)	(0.0026)	(0.0001)	(0.0035)	(0.0035)	(0.0001)
Log aversize	No	0.0120**	No	No	0.0319***	No
		(0.0047)			(0.0063)	
Log population	No	0.0061***	No	No	-0.0005	No
		(0.0022)			(0.0038)	
Log attract. index	No	-0.0411***	No	No	-0.0235***	No
		(0.0052)			(0.0082)	
Housing price (ratio)	No	0.0019***	No	No	0.0014***	No
		(0.0002)			(0.0004)	
Unemployment rate	No	-0.0043***	No	No	-0.0030**	No
		(0.0009)			(0.0014)	
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	No	Yes	Yes	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Constant	1.0472***	1.2230***		0.2211***	0.3260***	
	(0.0277)	(0.0382)		(0.0482)	(0.0592)	
Observations	12,380,994	12,358,595	12,380,994	4,870,496	4,866,528	4,870,496
R-squared	0.557	0.559		0.635	0.636	
RMSE	0.314	0.314		0.394	0.394	
σ_α			0.3750			0.4821
σ_ϵ			0.1424			0.1670
ρ			0.8739			0.8929
F test			17.28			19.75
Prob > F			0.0000			0.0000

Source: Computations of the author
Note: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1. For FE model, F test is the F test that all $\alpha_i = 0$

As expected, the private returns to education are higher for the most qualified workers, and an additional year of schooling of an individual yields an increase of about 3.3%

and 10% on wages of the less qualified and the most qualified workers, respectively. For the group of individuals who increased their schooling level over the period under study, the fixed-effects estimates show that the private returns to education for the most qualified workers are about five times larger than for the less qualified, i.e. an extra year of schooling of a qualified workers who has improved his education leads to an increase on wages of about 5% while this increase is about 0.8% for the less qualified workers, controlling to the unobserved individual characteristics.

It is found evidence that local human capital has a larger effect on wage of the most qualified workers, but for the less qualified workers the effect is also significant and positive. OLS estimates for the less qualified group reflect the sum of positive complementarity and spillover effects, and for most qualified group the external returns reflect the net impact of two opposing effects, and a positive estimates indicate that the positive spillover effect more than offsets the negative effect of the increase in their relative supply.

Note that, as in previously estimations, including the county controls slightly lowers the coefficients of local human capital. So, the estimated spillover effects indicate that an additional year in average education of the county increase the wage of most qualified workers by 4.4% and those of less qualified workers by 3.0%. By the other hand, an increase in the share of highly qualified workers of county by one percentage point increases wages of most qualified workers by 0.43% and of about 0.30% of the less qualified workers, or else an additional year of calibrated education in the county leads to a wage premium of about 2.5% for the most qualified workers and of about 1.6% for the less qualified workers. These results provide evidence on the existence of positive pure educational spillovers, and they are in line with results proposed in empirical literature (Acemoglu and Angrist, 2001; Moretti, 2004a; Isacsson, 2005; Dalmazzo and de Blasio, 2007; Liu, 2007; Heuermann, 2011; Rodríguez-Pose and Tselios, 2012). Concerning the other control variables, the sign and size of the coefficients are similar to the previously estimated results, underlining that there is lower discrimination against men for the the most qualified women than for the less qualified women.

Fixed-effects estimates reported in Table 3 again relate to workers who improved their education during the period 2002-2009 and it is confirmed the existence of positive and significant externalities of local human capital, whatever the measure of human capital used. So, the most qualified workers benefit about twice higher external returns to human capital than the less qualified workers[8], perhaps due to their greater ability to learn. The F test indicates that pooled OLS would be inappropriate and after performing the Hausman test is verified that the unobservable individual effects are correlated with the explanatory variables.

4.1.1 Instrumental variables results

Another alternative approach to deal with the problem of endogeneity is the instrumental variable (IV) estimation and the advantage of the IV method is that a valid instrument isolates the effect of exogenous changes in human capital levels on wages (Moretti, 2004a,b). In this analysis the presence of a Polytechnic Institute in a county is used as instrument to local human capital, following (Moretti, 2004a), and a second instrument it is proposed: the share of people who can read in the county in 1900, which is something new in empirical literature on external returns. Because in 1900 the number of counties is lower than 278, the number of observations in the sample decrease by about 7%, i.e. the sample used in the IV estimation contains of about 13.5 million of observations.

Individual education is also treated as endogenous variable and the change in compulsory school attainment in 1981, that established the compulsory education of 9 years, is used as instrument, which is a common instrument in empirical economics (see Angrist and Krueger, 1991; Acemoglu and Angrist, 2001).

The specification to estimate by IV method only includes the individual and firms controls, a set of dummies to sector of activity and time effects, using the different ways to measure human capital in county, average years of education, share of highly skilled workers and skill index. Note that, the control variables for the county are not used due to its collinearity with the instrument for the aggregate human capital, since the instrumental variables are included in order to control for the impact of unobservable time-region specific factors (Heuermann, 2011).

Since the panel structure is available, in order to complete this empirical analysis the instrumental variables fixed-effects estimation (IV-FE) is applied to circumvent the poten-

[8]This result also holds for the sample of workers who did not improve their education during the period 2002-2009. See Table 13 in Appendix A.1

tial endogeneity problem of human capital of the county. Because the sample used is too large, the software used is unable to perform the IV-FE estimation, thus with this method is only used an half of the sample, i.e. a sub-sample of more than 6 million of observations.

When testing for the potential endogeneity of the individual and local human capital variables, it appears that the hypothesis of exogeneity is rejected at the 1% and 5% levels. From the IV2SLS approach, Table 4 shows the regression results of local human capital equation.

Table 4: First-stage regression of local human capital variable

Human Capital Variable	Average Education IV	Average Education IV-FE	Share of HQ workers IV	Share of HQ workers IV-FE	Skill index IV	Skill index IV-FE
Polytechnic Institute	1.5838***	1.2550***	17.8921***	0.0090***	2.9675***	2.3578***
	(0.0241)	(0.0007)	(0.2766)	(0.0013)	(0.0422)	(0.0013)
R^2	0.6988		0.6894		0.6717	
F-test on excluded var.	2773***		2717***		2874***	
Share 1900	0.0660***	0.0623***	0.7422***	0.7031***	0.1204***	0.1141***
	(0.0003)	(0.0000)	(0.0040)	(0.0002)	(0.0007)	(0.0000)
R^2	0.8172		0.8142		0.7914	
F-test on excluded var.	19488***		19287***		16798***	
Observations	13,420,048	6,609,603	13,420,048	6,609,603	13,419,674	6,609,409

Source: Computations of the author
Note: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.

These results show that both instruments present a positive and significant impact on human capital of the county, which points out that the instruments used are relevant, i.e. it is supported that the share of people who can read in the county in 1900 and the presence of a Polytechnic Institute in a county have a positive and significant impact on aggregated human capital of the county. Moreover, both instruments have a pronounced weight in individual education (values not reported in Table 4).

Results reported in Table 4 also indicate that the R^2 of first stage regression of local human capital ranges between 0.6717 and 0.8172, and these values are higher than those that appear in the empirical literature. F-statistics on the exclusion instruments in the reduced form local human capital equation are quite high for both instruments, exceeding the rule-of-thumb value of 10 for non-weak instruments (see Bound et al., 1995 and Staiger and Stock, 1997). As such, it can not rejected the significance of the excluded instruments in the local human capital equation, in this way, the instruments are considered to be valid. Considering Table 4 and previous discussion (section 3), the instruments to local human capital proposed are likely to satisfy relevance and exogeneity properties.

Table 5 presents the estimation results from the OLS estimation and from the second stage of IV2SLS approach that uses both mentioned instruments for aggregated human capital, as well, the compulsory schooling of 9 years is used as instrument for individual education. In addition, in this Table is reported the instrumental variables fixed-effects estimates using the same instruments as in the IV approach. In all estimation methods used, the local human capital is measured by the three ways used so far.

Table 5: Instrumental variables estimates of external returns to human capital

	(1)	(2)	(3)	(4)	(5)
VARIABLES	**Dependent variable: Log of real hourly wage**				
	Measure of Human Capital used: Average education of the county				
	OLS	IV Polyt. Inst.	IV Share 1900	IV-FE Polyt. Inst	IV-FE Share 1900
Education	0.0484***	0.0656***	0.0648***	0.0321***	0.0308***
	(0.0005)	(0.0019)	(0.0019)	(0.0001)	(0.0001)
Average education	0.0455***	0.0217***	0.0330***	-0.0051***	0.0034***
	(0.0020)	(0.0049)	(0.0033)	(0.0003)	(0.0002)
Individual controls	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes
Constant	0.9430***	0.8717***	0.8057***		
	(0.0325)	(0.0404)	(0.0378)		
Observations	13,420,048	13,420,048	13,420,048	6,609,603	
R-squared	0.642	0.635	0.637		
RMSE	0.341	0.345	0.344		
	Measure of Human Capital used: Share of High Qualified workers in a county				
	OLS	IV Polyt. Inst.	IV Share 1900	IV-FE Polyt. Inst	IV-FE Share 1900
Education	0.0485***	0.0655***	0.0648***	0.0319***	0.0309***
	(0.0005)	(0.0019)	(0.0019)	(0.0001)	(0.0001)
Share of HQ workers	0.0041***	0.0019***	0.0029***	-0.0005***	0.0003***
	(0.0002)	(0.0004)	(0.0003)	(0.0000)	(0.0000)
Individual controls	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes
Constant	1.1868***	0.9889***	0.9845***		
	(0.0283)	(0.0344)	(0.0347)		
Observations	13,420,048	13,420,048	13,420,048	6,609,603	6,609,603
R-squared	0.642	0.635	0.637		
RMSE	0.341	0.345	0.344		
	Measure of Human Capital used: Skill index of county				
	OLS	IV Polyt. Inst.	IV Share 1900	IV-FE Polyt. Inst	IV-FE Share 1900
Education	0.0492***	0.0660***	0.0654***	0.0317***	0.0310***
	(0.0005)	(0.0019)	(0.0018)	(0.0001)	(0.0001)
Skill index (county)	0.0246***	0.0115***	0.0180***	-0.0027***	0.0018***
	(0.0011)	(0.0026)	(0.0018)	(0.0002)	(0.0001)
Individual controls	Yes	Yes	Yes	Yes	Yes

Continued on next page

Table 5 – continued from previous page

Occupation dummies	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes
Constant	1.1018***	0.9476***	0.9195***		
	(0.0291)	(0.0343)	(0.0346)		
Observations	13,419,674	13,419,674	13,419,674	6,609,409	6,609,409
R-squared	0.642	0.635	0.636		
RMSE	0.341	0.345	0.344		

Source: Computations of the author
Note: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.

The use of change of compulsory schooling as instrument leads to an increase in the estimated private returns to education from about 4.8% to around 6.5%-6.7%, while the estimates of educational external returns decrease in size (as in Acemoglu and Angrist, 2001; Iranzo and Peri, 2009), regardless of the measure of local human capital and the estimation method used. However, the estimates further indicate that statistically significant human capital externalities effects exist.

According to Moretti (2004a,b), the OLS bias is positive if positive shocks are associated with increases in aggregated human capital and individual income in a region. Thus, if variation on most qualified workers across counties is driven by unobserved demand for skilled workers factors, OLS estimates are biased upward. For the different measures of human capital used, the results IV are similar independently of the instrument applied, which in turn are similar to OLS results from the previous specification that includes control variables for characteristics of the county. Thus, for instance, an increase of the average education in a county by one year increase the wage of workers by around 2.2%-3.3%, or an increase in the share of most qualified workers of county by one percentage point raises wages by about 0.19%-0.29%. Finally, coefficients of county skill index also shows that an additional year of calibrated education in the county leads to a wage premium of about 1.2%-1.8%. The signs and magnitude of the remain variables are the expected.

The IV-FE estimates that are reported in columns 4 and 5 of Table 5 and they show that the private returns to education are higher than the fixed-effects estimates reported in Table 2, suggesting that measurement error has biased the fixed-effects estimates downward (Li et al., 2012). The IV-FE estimates of external returns to human capital have opposite signals when different instruments are used. The negative signal of the external returns to human capital when the presence of a Polytechnic Institute in a county is used as instrument is probably due the low variability of instrument, since there are Polytechnic Institute in only 6.5% of the mainland Portuguese counties, which distributed by about 13 million of observations yields a low standard-deviation of the instrument (0.2472). Thus, it is considered the share of people who can read in the county in 1900 as the only valid instrument and it is confirmed the existence of external returns to education in all measures of human capital.

Previous estimates of the external return to education are obtained by pooling all education groups together and therefore represent an average effect across education groups, that may not mean the existence of spillovers (Moretti, 2004a,b). Table 6 presents the OLS, IV and IV-FE estimates for two groups of workers defined above, the less qualified and the most qualified, to control to imperfect substitutability.

Table 6: IV results: Less qualified versus most qualified workers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	**Dependent variable: Log of real hourly wage**									
	Measure of Human Capital used: Average Education of the county									
	Less qualified workers					Most qualified workers				
	OLS	IV	IV	IV-FE	IV-FE	OLS	IV	IV	IV-FE	IV-FE
		Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900		Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900
Education	0.0332***	0.0519***	0.0512***	0.0171***	0.0164***	0.0994***	0.0814***	0.0802***	0.0569***	0.0535***
	(0.0006)	(0.0028)	(0.0028)	(0.0002)	(0.0002)	(0.0021)	(0.0030)	(0.0030)	(0.0004)	(0.0004)
Av. education	0.0429***	0.0222***	0.0296***	-0.0037***	0.0024***	0.0624***	0.0285***	0.0548***	-0.0049***	0.0107***
	(0.0020)	(0.0049)	(0.0033)	(0.0003)	(0.0003)	(0.0027)	(0.0063)	(0.0040)	(0.0006)	(0.0004)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Constant	0.9014***	0.8419***	0.8017***			0.0100***	0.5149***	0.3596***		
	(0.0328)	(0.0284)	(0.0381)			(0.0531)	(0.0657)	(0.0644)		
Observations	11,718,151	11,718,151	11,718,151	5,768,666	5,768,666	4,628,999	4,628,999	4,628,999	2,291,400	2,291,400
R-squared	0.5603	0.5513	0.5526			0.6365	0.6322	0.6345		
RMSE	0.342	0.3174	0.3169			0.395	0.397	0.396		
	Measure of Human Capital used: Share of High Qualified workers in a county									
	Less qualified workers					Most qualified workers				
	OLS	IV	IV	IV-FE	IV-FE	OLS	IV	IV	IV-FE	IV-FE
		Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900		Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900
Education	0.0333***	0.0518***	0.0511***	0.0170***	0.0165***	0.0994***	0.0814***	0.0800***	0.0567***	0.0540***
	(0.0006)	(0.0028)	(0.0028)	(0.0002)	(0.0002)	(0.0021)	(0.0030)	(0.0030)	(0.0004)	(0.0004)
Share of HQ	0.0039***	0.0020***	0.0026***	-0.0003***	0.0002***	0.0056***	0.0025***	0.0048***	-0.0004***	0.0009***
	(0.0002)	(0.0003)	(0.0004)	(0.0000)	(0.0000)	(0.0002)	(0.0005)	(0.0004)	(0.0001)	(0.0000)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Constant	1.1316***	0.9620***	1.3831***			0.3483***	0.6713***	0.6596***		
	(0.0328)	(0.0284)	(0.0381)			(0.0509)	(0.0628)	(0.0634)		
Observations	11,718,151	11,718,151	11,718,151	5,768,666	5,768,666	4,628,999	4,628,999	4,628,999	2,291,400	2,291,400
R-squared	0.5604	0.5513	0.5527			0.6366	0.6322	0.6346		
RMSE	0.314	0.317	0.317			0.395	0.397	0.396		
	Measure of Human Capital used: Skill Index of county									
	Less qualified workers					Most qualified workers				
	OLS	IV	IV	IV-FE	IV-FE	OLS	IV	IV	IV-FE	IV-FE
		Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900		Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900
Education	0.0339***	0.0523***	0.0517***	0.0169***	0.0166***	0.1003***	0.0816***	0.0805***	0.0564***	0.0545***
	(0.0006)	(0.0028)	(0.0027)	(0.0002)	(0.0002)	(0.0020)	(0.0030)	(0.0030)	(0.0004)	(0.0003)
Skill index	0.0231***	0.0161***	0.0117***	-0.0020***	0.0013***	0.0341***	0.0155***	0.0302***	-0.0027***	0.0060***
	(0.0011)	(0.0018)	(0.0026)	(0.0002)	(0.0001)	(0.0015)	(0.0034)	(0.0022)	(0.0003)	(0.0002)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Continued on next page

Table 6 – continued from previous page

Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Constant	1.0519***	0.9196***	0.9035***			0.2268***	0.6181***	0.5554***		
	(0.0281)	(0.0385)	(0.0388)			(0.0492)	(0.0611)	(0.0617)		
Observations	11,717,826	11,717,826	11,717,826	5,768,495	5,768,495	4,628,950	4,628,950	4,628,950	2,291,377	2,291,377
R-squared	0.5602	0.5510	0.5525			0.6363	0.6320	0.6343		
RMSE	0.314	0.318	0.317			0.395	0.397	0.396		

Source: Computations of the author
Note: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.

IV estimation results suggest that the most qualified workers receive a higher wage premium for working in counties with high human capital, however, less qualified workers also benefit from a significant wage premium, i.e. both groups of workers, the less qualified and the most qualified workers, obtain productivity gains in counties with the greatest concentration of human capital. This means the existence of productive externalities generated by the interaction with people more qualified. This result is in disagreement with one part of literature, for instance, Moretti (2004a) which suggest that unskilled workers receive higher wage premium than skilled workers, but it is in line with another part, for instance, Heuermann (2011) and Silva (2003) and the last author proposes that, after 5 years of being displaced, the wages of a college graduate displaced worker raise by more than 3% compared to a displaced worker with basic education if the average education of their county increase by one year. Note that, when the presence of a Polytechnic Institute in a county is used as instrument, the IV-FE estimates of the external returns to human capita present again a negative signal and IV estimates are similar in both groups of qualification workers, thus confirming the low variability of the instrument as discussed above. Therefore, the remaining analysis is only considered the other instrument proposed for the local human capital.

Thus, on one hand, Table 6 shows that an increase in the share of most qualified workers (or average education) in the county by one percentage point (one extra year) increases wages of most qualified workers by 0.48% (5.48%) compared to 0.56% (6.24%) in OLS regressions. By other hand, wage effects from human capital of county externalities are again smaller for the less qualified workers and an increase of one percentage point (one extra year) in share of most qualified workers (average education) raises wages of the less qualified by 0.26% (2.96%) compared to 0.39% (4.29%) in OLS regression. This finding implies that the existence of human capital spillovers cannot be reject for both qualification groups of workers. However, in both groups of workers the private returns to education are higher than the external returns to education, i.e. by IV estimation, to the most qualified workers an extra year in individual education leads to an increase on wages by about 8% and an increase in average education of the county leads to an increase of about 5.5%, and by IV-FE estimation these values are 5.35% and 1.07%, respectively. As respect to the less qualified workers, an extra year in individual education leads to an increase on wages by about 5.12% and an increase in average education of the county leads to an increase of about 2.96%, and by IV-FE estimation the corresponding values are 1.64% and 0.24%, respectively.

In addition, and as expected, the most qualified workers benefit from higher private returns to education. Note that, IV estimates of private returns to education are higher than the OLS estimates for the less qualified workers, however, for the most qualified workers the reverse is verified. This finding may depend on their individual characteristics, consisting on the ability of the workers with lower schooling to translate benefits from their education into higher individual productivity. According to the literature, the IV estimates are even more over-estimated than the corresponding OLS estimates due to the unobserved differences between both "treatment" and "comparison" groups implicit in IV approach (Card, 2001). Thus, this difference can be explained by the underlying heterogeneity in the returns to education, because factors, such as, changes in compulsory schooling are more likely to affect the educational choices of the individuals, which otherwise have low education. As such, this subgroup of workers will tend to benefit from marginal returns to education somewhat higher than the average returns to education of the population as a whole, and may lead the IV estimates of returns to education to exceed the corresponding OLS estimates (Card, 1999, 2001). By other words, this situation happens if the individuals increasing marginal discount rates, because taste or financial constraints factors, the marginal returns to education are higher for workers with lower education, indicating that they have greater ability (Card, 1999, 2001; Dickson, 2012).

The remained variables have the expected sign and magnitude, checking up again that the wage gap between men and women is higher in the less qualified than in the most qualified workers.

Finally, the Hausman-Taylor estimation is applied, however, the results are not reported, since the Hausman-Taylor estimates for the specification studied are inconsistent and the estimated results of the effect of education on individual wages results are in complete disagreement with the theory of human capital, as well, with results obtained for other estimation methods in this essay. The computed Hausman test, comparing the fixed-effects and Hausman-Taylor models, strongly rejects the null hypothesis, i.e. the fixed-effects estimator is consistent, while the Hausman-Taylor one is not. Possibly, the instruments used by the model are weak. It was attempted to apply a test of over-identifying restrictions for the Hausman-Taylor estimation, but it it seems that the STATA command (*xtoverid*) used does not work for a large sample.

4.2 Human capital externalities at regional and firm level

Different to the existent Portuguese literature and following Canton (2009) and Bauer and Vorell (2010), it is analysed both local and firm human capital effects on individual wages, allowing assess whether still exist local spillover effects of human capital and compare them with firm spillovers effects.

So, to go further in the investigation about the externalities of education, previous specification is augmented with a firm human capital term, which is measured in three different ways again: average years of education, share of highly skilled workers and skill index at the firm level. So the wage equation has the following way:

$$\log W_{irt} = X_{irt}\delta + \bar{S_{rt}}\eta_1 + \bar{S_{ft}}\eta_2 + Z_{rt}\beta + d_t + \mu_{irt} \quad (3)$$

with W_{irt} denoting the hourly wage of individual i in county r at time t; X_{irt} is a vector of observable characteristics of individual i in county r at time t, including gender, education, experience and tenure, and their corresponding squares, as well a set of dummies to occupation and activity sector, and a firm size control; $\bar{S_{rt}}$ represents regional human capital in the county r in time t; $\bar{S_{ft}}$ represents human capital in the firm f in time t; Z_{rt} is a vector of county characteristics; and d_t is a time fixed effects; and μ_{irt} is the error term to the unobservable characteristics and it is assumed to have the the basic assumptions of the classic regression model. The variables of interest are $\bar{S_{rt}}$ and $\bar{S_{ft}}$, and the estimated coefficients of these variable, η_1 and η_2, respectively, provide evidence of the existence of regional and firm-level spillover effects of education.

In order to meaningfully calculate human capital stocks at firm level, it is imposed that the number of workers in the firm must be lager than five. As in the previous sections, in this analysis robust standard errors are applied, clustered at the firm level, controlling for heteroscedasticity and possible serial correlation within firms.

Results from estimating equation (3) by pooled OLS are reported in columns 1–6 of the Table 7. However, pooled OLS estimates may be biased by time-invariant unobserved individual effects that may be correlated with $\bar{S_{rt}}$ and $\bar{S_{ft}}$. Therefore, columns 7–9 of the table report fixed-effects estimation.

Table 7: Human Capital externalities at local and firm levels: Pooled OLS and FE estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	**Dependent variable: Log of real hourly wage**								
	OLS						**Fixed-effects**		
Educ	0.0388***	0.0406***	0.04448***	0.0387***	0.0404***	0.0444***	0.0101***	0.0112***	0.0117***
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0006)	(0.0001)	(0.0001)	(0.0001)
Av. county education	0.0320***			0.0149***			0.0084***	0.0009***	
	(0.0020)			(0.0035)			(0.0001)		
Av. firm education	0.0397***			0.0397***			0.0128***		
	(0.0012)			(0.0012)			(0.0001)		
Share of HQ (county)		0.0030***			0.0020***			0.0009***	
		(0.0002)			(0.0003)			(0.0000)	
Share of HQ (firm)		0.0032***			0.0032***			0.0010***	
		(0.0001)			(0.0001)			(0.0000)	
Skill index (county)			0.0176***			0.0079***			0.0050***
			(0.0011)			(0.0021)			(0.0001)
Skill index (firm)			0.0320***			0.0318***			0.0106***
			(0.0012)			(0.0012)			(0.0001)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	No	No	No	Yes	Yes	Yes	No	No	No
Observations	12,091,700	12,091,700	11,397,056	12,072,259	12,072,259	11,379,109	12,091,700	12,091,700	11,397,056
R-squared	0.672	0.671	0.678	0.674	0.672	0.679			
RMSE	0.330	0.330	0.330	0.329	0.330	0.329			
σ_α							0.4543	0.4554	0.4620
σ_ϵ							0.1451	0.1451	0.1430
ρ							0.9074	0.9078	0.9126
F test							18.33	18.35	18.54
Prob > F							0.0000	0.0000	0.0000

Source: Computations of the author
Notes: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.
For FE model, F test is the F test that all $\alpha_i = 0$

When considering both aggregated and individual human capital measures jointly, the results indicate that human capital externalities exist both at the county level and at the firm level, regardless the human capital measure used. By comparing this pooled OLS estimation with results of the previous section (Table2), it is verified that the inclusion of aggregated human capital at the firm level makes the coefficient of aggregated human capital of the county slightly lower.

As in previous set of results, the magnitude of the coefficients of human capital at county level decrease when controls for county characteristics are included, but the coefficients remain the same at the firm level. Note that, the results suggest that the effect of firm's aggregated human capital on wages of the workers is equally strong as the aggregated human capital of the county, however, the size of the externalities are lightly higher on firm level than on county level. For instance, an increase of one year in the average county education yields an increase of about 1.49% on wages, while an increase of one year in the average firm education leads to an increase of about 3.97% on wages of the workers. By other hand, the estimated coefficients indicate that the wage of workers rise by about 0.32% when the share of the most skilled workers in a firm increases by one percentage point, while the increase is about 0.20% with an increase by one percentage point in the share of the most skilled workers of the county, and also show that an additional year of calibrated education in the firm leads to a external return of about 3.18% while this return is just about 0.79% by an additional year of calibrated education in the county. In accordance with literature, a possible interpretation for these findings is that human capital spillovers are strongest within the firm, as productive social interaction are probably to be more intense among colleagues than among people of a county (Canton, 2009). These results are in line with literature (see Canton, 2009). Furthermore, the pooled OLS estimates indicate that individual return to education and external returns to human capital of the firm are statistically identical to about 4%, while external return to human capital of the county is just about 1.49%. The estimation results indicate that pooled OLS delivers upward-biased effects of both local and firm human capital on individual wages, thus the estimated effects of aggregated human capital appear to be biased by unobserved time-invariant effects.

Controlling for time-invariant individual fixed-effects the results also provide evidence for the existence of significant and positive human capital externalities, for any measure of human capital used, but at lower magnitude. It is confirmed that the private return and external return at the firm level are statistically the same and higher than the external return at the county level for workers who have improved their schooling in period under study. The fixed-effects estimation is the most suitable model, since the F test indicates that there are significant unobserved individual effects, implying that pooled OLS would be inappropriate and the Hausman test's null hypothesis is rejected (p-value Hausman test is 0.0000).

These results and magnitude of human capital spillovers are not consistent with Portuguese literature known on this issue, the work of Martins and Jin (2010), which estimate the Mincer wage equations at firm-level using a matched employer-employee panel data from Portugal. The results found in this analysis are lower than the findings obtained by Martins and Jin (2010). Authors find evidence of firm-level returns of about 18% when the pooled OLS is used, much larger than standard estimates of private returns, and in the fixed-effects estimation, the estimated return falls considerably, but is still statistically significant and economically relevant, at 5%. However, the authors only consider the aggregated human capital at the firm level.

As previous noted, evidence of a stronger relationship between aggregated human capital and earnings does not necessarily, in general, imply positive externalities, due the imperfect substitution between skilled and unskilled workers which may induce such result. So, the results estimation for separate education groups are reported in Table 8.

Table 8: Spillovers at local and firm level: Less *vs* most qualified workers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	**Dependent variable: Log of real hourly wage**								
	Less qualified workers								
	Average education		Share of qualified workers		Skill index of county		Fixed-effects		
Education	0.0229***	0.0227***	0.0254***	0.0251***	0.0286***	0.0281***	0.0044***	0.0055***	0.0062***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0001)	(0.0001)	(0.0001)
Av. county	0.0320***	0.0155***					0.0056***		
education	(0.0020)	(0.0036)					(0.0001)		
Av. firm	0.0434***	0.0432***					0.0130***		
education	(0.0013)	(0.0013)					(0.0001)		
Share of HQ			0.0030***	0.0020***				0.0006***	
(county)			(0.0002)	(0.0004)				(0.0000)	
Share of HQ			0.0035***	0.0034***				0.0010***	
(firm)			(0.0001)	(0.0001)				(0.0000)	
Skill index					0.0171***	0.0074***			0.0036***
(county)					(0.0012)	(0.0022)			(0.0001)
Skill index					0.0350***	0.0348***			0.0097***
(firm)					(0.0012)	(0.0012)			(0.0001)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	No	Yes	No	Yes	No	Yes	No	No	No
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Constant	1.0342***	1.2376***	1.4093***	1.5671***	1.2645***	1.3831***			
	(0.0328)	(0.0381)	(0.0284)	(0.0452)	(0.0312)	(0.0441)			
Observations	12,381,330	12,358,931	12,381,330	12,358,931	12,380,994	12,358,595	10,444,174	10,444,174	9,890,349
σ_α							0.3748	0.3758	0.3794
σ_ϵ							0.1392	0.1392	0.1379
ρ							0.8787	0.8793	0.8833
F test							16.91	16.94	17.06
Prob > F							0.0000	0.0000	0.0000
VARIABLES	**Dependent variable: Log of real hourly wage**								
	Most qualified workers								
	Average education		Share of qualified workers		Skill index of county		Fixed-effects		
Education	0.0906***	0.0907***	0.0950***	0.0951***	0.0995***	0.0994***	0.0484***	0.0502***	0.0505***

Continued on next page

Table 8 – continued from previous page

	Average education		Share of qualified workers		Skill index of county		Fixed-effects		
	(0.0018)	(0.0019)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0003)	(0.0003)	(0.0003)
Av. county education	0.0337***	0.0151***					0.0066***		
	(0.0028)	(0.0057)					(0.0003)		
Av. firm education	0.0500***	0.0507***					0.0149***		
	(0.0016)	(0.0016)					(0.0001)		
Share of HQ (county)			0.0034***	0.0023***				0.0009***	
			(0.0003)	(0.0006)				(0.0000)	
Share of HQ (firm)			0.0037***	0.0038***				0.0010***	
			(0.0001)	(0.0001)				(0.0000)	
Skill index (county)					0.0214***	0.0141***			0.0054***
					(0.0017)	(0.0037)			(0.0001)
Skill index (firm)					0.0387***	0.0387***			0.0099***
					(0.0017)	(0.0017)			(0.0001)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	No	Yes	No	Yes	No	Yes	No	No	No
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Constant	0.0398***	0.2138***	0.4316***	0.5398***	0.2839***	0.3803***			
	(0.0516)	(0.0596)	(0.0508)	(0.0656)	(0.0551)	(0.0689)			
Observations	4,870,546	4,866,578	4,870,546	4,866,578	4,870,496	4,866,528	4,283,954	4,283,954	4,010,356
σ_α							0.4761	0.4777	0.4823
σ_ϵ							0.1602	0.1602	0.1574
ρ							0.8984	0.8989	0.9036
F test							19.44	19.58	19.91
Prob > F							0.0000	0.0000	0.0000

Source: Computations of the author
Note: Robust standard errors in parentheses, clustered at the firm level.*** p<0.01, ** p<0.05, * p<0.1.
For FE model, F test is the F test that all $\alpha_i = 0$

Table 8 presents that the estimated coefficient of all human capital variables appear to be positive and statistical significant at 1% level. Results confirm again that workers with higher education benefit from higher private returns and magnitude of the human capital coefficients at the county level decrease when controls for county characteristics are included, but the coefficients remain similar at the firm level. Furthermore, it appears again that the aggregated human capital outside the firm exerts a significant influence on wages of both groups of workers and the externalities of firm human capital are higher than at county level. For instance, pooled OLS estimation indicates that an extra year in average education of the firm yields an increase on wages of the most qualified workers (less qualified) of about 5% (4.3%), while the corresponding value at county level is of about 1.5% for both groups of workers. Also presents that an increase of the share of high skilled in a firm of one percentage point increases the wages of the most qualified (less qualified) by about 0.37% (0.34%), the corresponding estimate at county level is about 0.23% (0.20%).

In addition, an additional year of calibrated education in the firm raises the wages of the most qualified workers (less qualified) by about 3.9% (3.5%) whereas the corresponding estimate at county level is of about 1.41% (0.74%). Within each level of aggregation the estimated coefficients are statistically identical for both worker skill groups.

Moreover, controlling for unobserved time-invariant individual effects, the estimated effect of the aggregated human capital on wages is lower than the effect obtained by pooled OLS, but still positive and significant at 1% level. It is confirmed that external returns at firm level are higher than external returns at county level, and in each level of aggregation the spillover magnitudes are statistically identical for both groups of workers.

These findings are partially in agreement with international literature, since in the literature the results are contradictory.

On one hand, Canton (2009) finds an important relationship between individual wages and both regional and firm human capital indicators, when the fraction of high-skilled workers is used as measure of aggregated human capital. Author suggests that magnitude of human spillovers at the firm level are higher than the human capital spillovers at the county and if aggregated human capital is measured by the average education of region/firm, no evidence is found for human capital spillovers at regional level and at firm level it is found an external return by about 1.3%. However, when controlling to imperfect substitution between workers no evidence is found for human capital spillovers for workers whose educational level is Higher Education and the lower the educational level greater the human capital spillovers at firm level, since at regional level the coefficient are no statistically significant.

Opposed results are found by Bauer and Vorell (2010). The author, using the regional share of high-skilled workers as measure of human capital and pooled OLS estimation, propose evidence of human capital spillovers only to the high-skilled workers (just at 10% level) and external return to regional human capital (0.310) is higher than external return to human capital at firm level (0.186). Controlling to individual fixed-effects, the authors find evidence of human capital externalities at regional level and for low-skilled workers (0.129), but they find external returns to human capital only at firm level and for high-skilled workers (0.324). Relative to Martins and Jin (2010) presents evidence of significant wage spillovers to less-educated workers, since their wages increase by 2% to 3% per extra year of education of workers in their firm, but group of educated workers does not seem to benefit from such spillovers.

4.3 Inter-regional Human Capital externalities

Over a long period of time, variables related to the location were not considered relevant to the explanation of many phenomena resulting from economic dynamic. However, with the emergence of adequate data and the development of techniques of spatial econometrics, the influence through effects spatial spillovers of neighbouring regions in the development of a given region has lately been considered in the empirical literature (Anselin, 2003).

However, the use of locational data may lead to spatial dependence between observations, which is an inherent characteristic to the representation of data through territorial subdivisions (Anselin, 1988). The spatial dependence can be defined as a lack of independence between observations (Anselin, 1988; LeSage, 2008; LeSage and Pace, 2009). A factor associated with the spatial dependence is the possible existence of externalities between observations, for instance, regions with high levels of human capital tend to have positive spillover to other (Anselin, 1988; LeSage, 2008). This assumption is directly related to the so-called first law of geography enunciated by Tobler (1970) "Everything is related to everything else, but near things are more related than distant things". Thus, the presence of spatial dependence violates the basic assumptions of Gauss-Markov generally used in the regression models, namely, it is assumed that observations/regions are independent and identically distributed, i.e. statistically independent observations imply that $E(\epsilon_i \epsilon_j) = E(\epsilon_i)E(\epsilon_i) = 0$. But, in contrast, spatial dependence reflects a situation where values observed at one region, observation i, depend on the values of neighbouring observation at nearby regions (LeSage and Pace, 2009). A tool created in econometrics to deal with spatial dependence is the spatial weight matrix, usually called W, which summarizes the spatial correlation structure (Anselin, 1988).

While literature on inter-regional knowledge spillovers has emerged, namely, the the contribution of inter-regional spillovers in the aggregate regional wage (Rodríguez-Pose and Tselios, 2012), the research to measure the relationship between individual wages and inter-regional educational externalities seems to be made only by Rodríguez-Pose and Tselios (2012). The authors use microeconomic data from the European Community Household

Panel, and resort to spatial economic analysis, across the regions of the European Union, for the purpose to determine to what extent differences in individual earnings are the result of the educational attainment of the individual, the educational attainment of the members of the individual household, the educational endowment of the region where the individual lives, or the educational endowment of the neighbouring regions. The results show that, in addition to the expected positive private returns, place-based regional and supra-regional educational educational externalities generate significant benefits for workers (Rodríguez-Pose and Tselios, 2012).

Therefore, following Rodríguez-Pose and Tselios (2012), in this section the goal is to determine evidence if there is presence or absence of supra-regional externalities of human capital on individual wages in Portugal. This external return will occur if, for instance, an improvement in educational level of the people who works in a county contributes to an increase in the individual wages of an individual who works in a neighbouring counties, i.e. workers with the similar educational characteristics working in different counties may have different wages (Rodríguez-Pose and Tselios, 2012). So, the model to estimate is given by:

$$\log W_{irt} = X_{irt}\beta + \bar{S_{rt}}\eta_1 + [W\bar{S}_t]_r\eta_2 + Z_{rt}\delta_1 + [WZ_t]_r\delta_2 + d_t + \mu_{irt} \tag{4}$$

with W_{irt} denoting the hourly wage of individual i in county r at time t; X_{irt} is a vector of observable characteristics of individual i in county r at time t, as previously, including gender, education, experience and tenure, and their corresponding squares, as well a set of dummies to occupation and activity sector, and a firm size control; $\bar{S_{rt}}$ represents regional human capital in the county r in time t; and $[W\bar{S}_t]_r$ represents the average education of neighbouring counties of county r at time t; Z_{rt} is a vector of county characteristics; $[WZ_t]_r$ control variables of neighbouring counties; and d_t is a time fixed effects; and μ_{irt} is the error term to the unobservable characteristics and it is assumed to have the the basic assumptions of the classic regression model. The variables of interest are $\bar{S_{rt}}$ and $[W\bar{S}_t]_r$, and the estimated coefficients of these variables, η_1 and η_2, respectively. They provide an indication of the existence of regional and inter-regional externalities of human capital.

These effects may not reflect the true impact of external returns to human capital, but county and neighbouring county characteristics that are themselves correlated with the educational attainment a county and broader geographic levels, respectively. With aim to eliminate these potential effects, a vector of individual-specific, county-specific and inter-regional-specific control variables are included (Rodríguez-Pose and Tselios, 2012). Matrix W is the spatial weight matrix and quantifies the connections between regions, i.e. the matrix W aims to capture the possible spatial correlation structure presented in the data, and your choice is one of the most difficult and controversial methodological issues, since the incorrect choice of W may lead to inconsistent estimates, affecting the interpretation of the parameters of wage equation (Anselin, 1988; LeSage, 2008). Thus, the choice of spatial weight matrix is supported by theoretical considerations.

By means of the spatial weights matrix W, a neighbour set is specified for each location, i.e. W is a 278×278 matrix where each element w_{ij} expresses the proximity between spatial object i (in the row of the matrix) and j (column) and can be interpreted as the strength of potential interaction between units i and j. So it can be interpreted as the presence and strength of a link between nodes (the observations) in a network representation that matches the spatial weights structure (Anselin, 1988; Anselin et al., 2008; Arbia and Baltagi, 2009).

The specification of W is in general arbitrary, based on some measures of distance between units. In this analysis, it is is assumed that a county is not considered its own neighbour, i.e. the diagonal elements of matrix W are equal to zero, $w_{ii} = 0$. According to the literature, the choice of specification of spatial weights is generally conducted by geographic criteria (Anselin et al., 2008). Thus, the assignment of weights w_{ij} is based on information on the contiguity among mainland Portuguese counties, and assigning $w_{ij} = 1$ when county i and j share a common border or vertex (contiguity matrix of queen type) and $w_{ij} = 0$ otherwise, as such, it is a symmetric matrix. By other words, as to spatial dependence, it is assumed that neighbouring counties, contiguous, present a greater degree of dependence than the others.

Following most applied literature on spatial econometrics, for computational simplicity and to aid in the interpretation of the spatial variables, the spatial weights are standardized, such that the elements in each row sum to 1, i.e. $w_{ij}^S = \frac{w_{ij}}{\sum_j w_{ij}}$ (Anselin et al., 2008), and although the original matrix W be symmetrical, the row-normalized matrix may not be symmetric. In this sense, $[W\bar{S}_t]_r = \sum_{j \neq r} w_{ir}\bar{S_{jt}}$, which is a linear combination of the observations of all counties r, as mentioned above, represents the average education of neighbouring counties of county r at time t and it is called spatial lag variable.

The data used in this section is the previously used. Table 9 reports the pooled OLS results of the Mincerian equation 4, and such as in the previous sections, regional human capital is measured by three ways: average years of education, share of highly skilled workers and the skill index in the county. These three measures of human capital are also used to determine the human capital endowment of the neighbouring counties. Furthermore, in the empirical analysis is exploited the structure panel using fixed-effects estimation, as in previous sections.

Table 9: Regional and Inter-regional externalities: Pooled OLS and FE estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	**Dependent variable: Log of real hourly wage**								
	Pooled OLS						**Fixed-effects**		
Educ	0.0484***	0.0485***	0.0491***	0.0487***	0.0487***	0.0491***	0.0130***	0.0124***	0.0136***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0001)	(0.0001)	(0.0001)
Av. county educ	0.0455***			0.0207***			0.0121***		
	(0.0019)			(0.0029)			(0.0001)		
Education neighbours	0.0016			0.0859***			0.0022***		
	(0.0015)			(0.0115)			(0.0001)		
Share of HQ (county)		0.0042***			0.0021***			0.0013***	
		(0.0002)			(0.0003)			(0.0000)	
Share of HQ (neighbours)		0.0011***			0.0090***			0.0021***	
		(0.0004)			(0.0013)			(0.0000)	
Skill index (county)			0.0250***			0.0118***			0.0068***
			(0.0011)			(0.0017)			(0.0001)
Skill index (neighbours)			0.0024			0.0564***			0.0031***
			(0.0015)			(0.0090)			(0.0001)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	No	No	No	Yes	Yes	Yes	No	No	No
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Neighbouring controls	No	No	No	Yes	Yes	Yes	No	No	No
Observations	14,166,402	14,166,402	12,254,726	12,586,670	12,586,670	10,675,032	14,166,402	14,166,402	14,166,016
R-squared	0.640	0.640	0.642	0.639	0.639	0.642			
RMSE	0.341	0.341	0.340	0.342	0.342	0.342			
σ_α							0.4513	0.4486	0.4531
σ_ϵ							0.1491	0.1491	0.1491
ρ							0.9015	0.9005	0.9522
F test							18.84	18.80	18.77
Prob > F							0.0000	0.0000	0.0000

Source: Computations of the author
Notes: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.
For FE model, F test is the F test that all $\alpha_i = 0$

As in previous sections and in line with the human capital theory, the internal return to education is of about 5% for each additional year in school, by pooled OLS estimation. Specifications in which control variables to characteristics of county and to characteristics of the neighbouring counties are included (columns 4-6), the dimension of the coefficients for human capital externalities at county level decrease, however, remain positive and significant. And, as opposed, the coefficients for human capital externalities of the neighbouring counties increase, and become significant for all measures of human capital . So, the results show a positive and significant effect of externalities of both levels, at county level and extra-county level, on individual wages, supporting the importance of the influence through effects educational spatial spillovers of neighbouring regions.

However, OLS estimates may be biased due the endogeneity of human capital variables. As such, individual fixed-effects estimation is use and the results are reported in columns 7-9 of the Table 9. Results confirm evidence for the existence of positive human capital externalities for workers, at different levels of aggregation.

Thus, it seems to be possible to assert that the wage of any worker are affected by the conjunction of the individual education, the human capital level of the county where the individual works and the human capital level of the neighbouring counties, but the individual educational level is the factor with more relevant weight.

Note that, if aggregated human capital is defined in terms either average educational attainment or skill index, the effect of externalities at county level is higher than the effect of externalities at extra-county level, but if aggregated human capital is defined by the share of qualified workers the opposite is verified. Thus, human capital spillovers at county level matter for individual earnings, but also inter-regional geographical spillovers generate benefits for workers, and this results is also proposed by Rodríguez-Pose and Tselios (2012). Therefore, an individual that works in a county with a high human capital level, whose neighbouring counties have similar levels of human capital, would tend to have higher wages than an individual with similar characteristics, but working in a county with lower human capital level, surrounded by other counties with similar human capital level (Rodríguez-Pose and Tselios, 2012).

Controlling again to the imperfect substitution between workers with different educational levels, which may induced the previous results, the estimation for separate education groups is applied and results are reported in Table 10.

Table 10 shows that the results still provide evidence on the existence of pure positive spillover effects for both qualifications groups of workers, using both estimation methods. However, the magnitude of the spillovers effects is lower when the individual fixed-effects estimation is used.

The F test indicates that there are significant unobserved individual effects, implying that pooled OLS would be inappropriate. And, the Hausman test's null hypothesis, that

Table 10: Regional and Inter-regional externalities: Less *vs* most qualified workers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	**Dependent variable: Log of real hourly wage**								
	Less qualified workers								
	Pooled OLS						**Fixed-efects**		
Education	0.0333***	0.0334***	0.0340***	0.0331***	0.0331***	0.0335***	0.0078***	0.0076***	0.0082***
	(0.0001)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0001)	(0.0001)	(0.0001)
Av. county education	0.0433***			0.0236***			0.0084***		
	(0.0019)			(0.0030)			(0.0001)		
Education neighbours	0.0021			0.0735***			0.0006***		
	(0.0014)			(0.0110)			(0.0001)		
Share of HQ (county)		0.0039***			0.0022***			0.0010***	
		(0.0002)			(0.0003)			(0.0000)	
Share of HQ (neighbours)		0.0011***			0.0077***			0.0012***	
		(0.0004)			(0.0012)			(0.0000)	
Skill index (county)			0.0233***			0.0128***			0.0047***
			(0.0011)			(0.0016)			(0.0001)
Skill index (neighbours)			0.0029**			0.0413***			0.0012***
			(0.0014)			(0.0078)			(0.0001)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	No	No	No	Yes	Yes	Yes	No	No	No
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Neighbouring controls	No	No	No	Yes	Yes	Yes	No	No	No
Observations	12,381,330	12,381,330	12,380,994	10,950,270	10,950,270	10,949,966	12,381,330	12,381,330	12,380,994
R-squared	0.5571	0.5572	0.5571	0.5567	0.5568	0.5566			
RMSE	0.314	0.314	0.314	0.315	0.315	0.315			
σ_α							0.3735	0.3716	0.3747
σ_ϵ							0.1424	0.1424	0.1424
ρ							0.8730	0.8719	0.8738
F test							17.33	17.30	17.28
Prob > F							0.0000	0.0000	0.0000
	Most qualified workers								
	Pooled OLS						**Individual fixed–efects**		
Education	0.0997***	0.0997***	0.1009***	0.0998***	0.0998***	0.1006***	0.0512***	0.0496***	0.0511***
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0002)	(0.0002)	(0.0003)
Av. county education	0.0614***			0.0249***			0.0135***		
	(0.0027)			(0.0051)			(0.0002)		
Education neighbours	0.0024			0.1486***			0.0033***		
	(0.0027)			(0.0140)			(0.0002)		
Share of HQ (county)		0.0055***			0.0020***			0.0015***	
		(0.0002)			(0.0005)			(0.0000)	
Share of HQ (neighbours)		0.0023***			0.0163***			0.0018***	
		(0.0004)			(0.0015)			(0.0001)	
Skill index (county)			0.0332***			0.0120***			0.0088***
			(0.0015)			(0.0031)			(0.0001)
Skill index (neighbours)			0.0041**			0.0920***			0.0061***
			(0.0028)			(0.0097)			(0.0002)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	No	No	No	Yes	Yes	Yes	No	No	No
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Neighbouring controls	No	No	No	Yes	Yes	Yes	No	No	No
Observations	4,870,546	4,870,546	4,393,982	4,427,817	4,427,817	3,951,260	4,870,546	4,870,546	4,393,892
R-squared	0.6353	0.6355	0.6346	0.6369	0.6370	0.6360			
RMSE	0.394	0.394	0.395	0.394	0.394	0.394			
σ_α							0.4835	0.4850	0.4857
σ_ϵ							0.1670	0.1670	0.1654
ρ							0.8934	0.8940	0.8961
F test							19.83	19.76	18.68
Prob > F							0.0000	0.0000	0.0000

Source: Computations of the author
Notes: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1. For FE model, F test is the F test that all $\alpha_i = 0$

the random-effects estimator is consistent, is rejected with a $p - value = 0$, therefore, the unobservable individual effects are correlated with the explanatory variables which makes the fixed effects estimation the most suitable model.

So, the effect of human externalities is higher for the most qualified workers than for the less qualified workers. Comparing the spillover effects of the aggregated human capital at different aggregation levels, if aggregated human capital is defined in terms either average educational attainment or skill index, the effect of externalities at county level is higher than the effect of externalities at extra-county level in both groups of workers, but if aggregated human capital is defined by the share of qualified workers the spillover effects are statistically identical.

Thus, it seems evident that the regression results show that the role of educational attainment at the three levels of analysis - individual, regional and inter-regional - matters for the wages of both the less and the most qualified workers. The results for both qualification groups of workers can not be compared with other studies, since the only known research on the study of the relationship between the potential effect of inter-regional educational spillovers and individual wages is the work of Rodríguez-Pose and Tselios (2012) and the authors did not analyse the role of imperfect substitutability between workers with different educational levels.

5 Concluding remarks

In this essay it is examined the existence and size of human capital spillovers at different levels of analysis, in the mainland Portuguese counties over the period 2002–2009. The estimation framework is the Mincerian wage regression augmented with the local human capital variable applied to a matched employer-employee panel set and instrumenting both individual and aggregated human capital, using the compulsory schooling of nine years of education, and the presence of a Polytechnic Institute in a county and the share of people who can read in the count based on the data of Census 1900, respectively.

In this analysis, three forms of aggregated human capital are distinguished, namely average years of education in the county, share of highly qualified workers in the county workforce and finally, a skill index in the county, because different forms of measure of regional human capital can be characterized by different spillovers.

Initially, the external returns to education are estimated by pooling all education groups, but subsequently two groups were defined to take account of the imperfect substitutability. Furthermore, since there are some methodological issues that must be addressed, different estimation methods are applied, i.e. in addition to the OLS estimation, the fixed-effects, the instrumental variables and instrumental variables fixed-effects are used.

The results suggest a positive spillover effect of human capital of the county on the wages of workers. For instance, an extra year in the average education of the county increase the wages by about 3%, on average, and this finding is in line with empirical literature (Moretti, 2004a,b; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009; Heuermann, 2011; Rodríguez-Pose and Tselios, 2012), whereas private amount roughly 4.8%–6.7%. The external return to education when regional human capital is defined in terms of average educational attainment is about 3.8% and of about 2.7% for the most qualified and less qualified workers, respectively, and an increase in the regional share of high qualified workers (12 or more years of schooling) of one percentage point is associated with a 0.33% and 0.23% increase in individual wages of the most qualified and the less qualified workers, respectively. So, it is found that the effects of the aggregated human capital on wages increase with the individual education in Portugal (as in Heuermann, 2011).

Different to the existent Portuguese literature and following Canton (2009) and Bauer and Vorell (2010), it is analysed the effects of both local human capital and human capital in a firm on individual wages, allowing assess whether still exist local spillover effects of human capital and compare them with firm spillovers effects. The inclusion of aggregated human capital in a firm level decreases the coefficient of aggregated human capital at county level, but remain positive and significant. The estimated spillover effects indicate that an increase of the share of high qualified workers in a firm of one percentage point increases the wages of the most qualified and less qualified workers in this firm by 0.38% and 0.34%, respectively, and an extra year in average education of the firm yields an increase on wages of the most qualified workers of about 5.1% and of about 4.3% on wages of the less qualified workers.

Finally, this analysis assesses the relationship between geographical (regional and inter-regional) human capital externalities across the mainland Portuguese counties and individual wages, since regions with high levels of human capital tend to have positive spillover to other (Anselin, 1988; LeSage, 2008). The results indicate that beyond the individual education and the human capital level of the county where the individual workers, the human capital level of the neighbouring counties also matters. In addition, both the less and the most qualified workers benefit by these external returns to human capital at the three aggregation levels.

In summary, individual wages are determined not solely by an individual's own education - private returns to education - but but also the knowledge gained through interaction with other workers and counties - external returns to human capital - comprising firm, regional and inter-regional externalities.

On one hand, the positive human capital spillovers that arise from average educational attainment might influence policies to combat early school leaving, to the extent that an increase in the school leaving can negatively influence the regional average education and, by other hand, positive spillovers associated with the presence of high-skilled workers more attention might influence the higher education policies.

These findings confirm the existence of external effects of human capital and according to Acemoglu and Angrist (2001) external returns to education of this magnitude are sufficient to justify significant public subsidies for education. Thus, according with findings obtained, the local development policy based on investment in local human capital would generate positive effects beyond the private return of the workers themselves. Thus, high private returns added to these spillovers human capital provide justification for public support of education. Furthermore, this public support of education leads to greater equality in education, between poor and wealthy families, and hence in individual wages and, therefore, increased wage equality tends to increase social cohesion, providing a further efficiency argument in favour of public support of education.

References

Acemoglu, D. and Angrist, J. (2001). *How Large are the Social Returns to Education: Evidence from Compulsory Schooling Laws.* MA: The MIT Press, Cambridge, 9–59.

Angrist, J. (1995). "The economic returns to schooling in the west bank and gaza strip."

American Economic Review, 85, 1065–1087.

Angrist, J. and Krueger, A. (1991). "Does compulsory school attendance affect schooling and earnings?." *Quarterly Journal of Economics*, 106, 979–1014.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, Dordrecht.

Anselin, L. (2003). "Spatial econometrics." *International Regional Science Review*, 26(2), 7–152.

Anselin, L., Gallo, J. L., and Jayet, H. (2008). *Spatial panel econometrics.* Springer-Verlag, Berlin, Chapter Chapter 19, 625–660.

Arbia, G. and Baltagi, B. (2009). *Spatial Econometrics: Methods and Applications Studies in Empirical Economics.* Physica-Verlag A Springer Company, Heidelberg.

Aziariadis, C. and Drazen, A. (1990). "Threshold externalities in economic development." *Quarterly Journal of Economics*, 105, 501–526.

Baltagi, B. (2008). *Econometric Analysis of Panel Data - 4th edition.* John Wiley and Sons Ltd, Chichester.

Baltagi, B., Bresson, G., and Pirotte, A. (2003). "Fixed effects, random effects or hausman-taylor? a pretest estimator." *Economics Letters*, 79, 361–369.

Bauer, T. and Vorell, M. (2010). "External effects of education: Human capital spillovers in regions and firms." *Report no.*, Ruhr Economic Papers 195.

Black, D. and Henderson, V. (1999). "A theory of urban growth." *Journal of Political Economy*, 107(2), 252–284.

Bound, J., Jaeger, D., and Baker, R. (1995). "Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak." *Journal of the American Statistical Association*, 90, 443–450.

Canton, E. (2009). "Human capital externalities and proximity: evidence from repeated cross-sectional data." *De Economist*, 157 (1), 79–105.

Card, D. (1995). "Earnings, schooling, and ability revisited." *Research in Labor Economics*, 14, 23–48.

Card, D. (1999). *Handbook of Labor Economics*. Elsevier, Amsterdam, Chapter The Causal Effect of Education on Earnings, 1801–1864.

Card, D. (2001). "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica*, 69(5), 1127–1160.

Cardoso, A. (1991). "Desigualdades salariais em portugal : uma perspectiva regional, 1982-1989." M.S. thesis, Universidade Técnica de Lisboa. Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa. Instituto Superior de Economia e Gestão.

Ciccone, A. and Peri, G. (2006). "Identifying human-capital externalities: Theory with applications." *Review of Economic Studies*, 73(2), 381–412.

Combes, P., Duranton, G., and Gobillon, L. (2011). "The identification of agglomeration economies." *Journal of Economic Geography*, 11(2), 253–266.

Dalmazzo, A. and de Blasio, G. (2007). "Social returns to education in italian local labor markets." *Annals of Regional Science*, 41(1), 51–69.

Duranton, G. (2006). *A Companion to Urban Economics*. Blackwell Publishing Ltd, Oxford, Chapter Human capital externalities in cities: identification and policy issues, 24–39.

Garcia-Fontes, W. and Hidalgo, M. (2009). "Estimating human capital externalities: The case of spanish regions." *Report no.*, Working Paper Econ 09.17.

Grácio, S. (1998). *Ensinos Técnicos e Política em Portugal 1910/1990*. Lisboa, Chapter 11, 12, 175 – 202.

Greene, W. (2012). *Econometric Analysis-7th edition*. Pearson Education.

Heckman, J. (2000). "Policies to foster human capital." *Research in Economics*, 51(1), 3–56.

Heckman, J. and Klenow, P. (1998). *Policies to Promote Capital Formation.* Stanford, CA: Hoover Institution, Chapter Human Capital Policy.

Heuermann, D. (2011). "Human capital externalities in western germany." *Spatial Economic Analysis*, 6(2), 139–165.

Hidalgo, M. A. (2009). "A demand and supply analysis of the spanish education premium in the 1980s and 1990s." *Report no.*, Discussion paper, Forthcoming Revista de Economía Aplicada.

Iranzo, S. and Peri, G. (2009). "Schooling externalities, technology, and productivity: Theory and evidence from u.s. states." *The Review of Economics and Statistics*, 91 (2), 420–431.

Isacsson, G. (2005). "External effects of education on earnings: Swedish evidence using matched employee-establishment data." *Report no.*, Institute for Labour Market Policy Evaluation (IFAU) Working Paper 10/2005, IFAU.

Katz, L. F. and Murphy, K. M. (1992). "Changes in relative wages,1963-1987: Supply and demand factors." *The Quarterly Journal of Economics*, 107(1), 35–78.

Lange, F. and Topel, R. (2006). *Handbook of the Economics of Education*, Vol. Vol I. Elsevier, Amsterdam, Chapter The Social Value of Education and Human Capital.

LeSage, J. (2008). "An introduction to spatial econometrics." *REVUE D' ECONOMIE INDUSTRIELLE*, 123, 19–44.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics.* Taylor and Francis Group, London.

Li, H., Liu, P., and Zhang, J. (2012). "Estimating returns to education using twins in urban china." *Journal of Development Economics*, 97, 494–504.

Liu, Z. (2007). "The external returns to education: Evidence from chinese cities." *Journal of Urban Economics*, 61, 542–564.

Lucas, R. J. (1988). "On the mechanics of economic development." *Journal of Monetary Economics*, 22(1), 3–42.

Martins, P. and Jin, J. (2010). "Firm-level social returns to education." *Journal of Population Economics*, 23, 539–558.

Moretti, E. (2004a). "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *Journal of Econometrics*, 121(1-2), 175–212.

Moretti, E. (2004b). *Handbook of Regional and Urban Economics*. Elsevier, Amsterdam, Chapter Human Capital Externalities in Cities, 2243–2291.

Moretti, E. (2004c). "Workers' education, spillovers, and productivity: Evidence from plant-level production functions." *American Economic Review*, 94(3), 656–690.

Muravyev, A. (2008). "Human capital externalities: Evidence from the transition economy of russia." *Economics of Transition*, 16(3), 415–443.

Portela, M. (2001). "Measuring skill: a multi-dimensional index." *Economics Letters*, 72, 27–32.

Rauch, J. (1993). "Productivity gains from geographic concentration of human capital: Evidence from the cities." *Journal of Urban Economics*, 34(3), 380–400.

Roback, J. (1982). "Wages, rents, and the quality of life." *The Journal of Political Economy*, 90(6), 1257–1278.

Rodríguez-Pose, A. and Tselios, V. (2012). "Individual earnings and educational externalities in the european union." *Regional Studies*, 46(1), 39–57.

Rudd, J. (2000). "Empirical evidence on human capital spillovers." *Report no.*, Finance and Economics Discussion Series 2000-46, Board of Governors of the Federal Reserve System (U.S.).

Silva, J. (2003). "Local human capital externalities or sorting? evidence from a displaced workers sample." *Report no.*, NIPE, Working Paper 9/2003.

Staiger, D. and Stock, J. (1997). "Instrumental variables regression with weak instruments." *Econometrica*, 65(3), 557–586.

Teulings, C. and Vieira, J. (2004). "Urban versus rural return to human capital in portugal: a cookbook recipe for applying assignment models." *Labour Economics*, 18, 265–291.

Verbeek, M. (2008). *A Guide to Modern Econometric- 3rd edition*. J. Wiley and Sons, Ltd, Rotterdam, Chapter Endogeneity, Instrumental variables and GMM, 129–167.

Vieira, J., Couto, J., and Tiago, M. (2005). "Inter-regional wage dispersion in portugal." *Report no.*, Institute for the Study of Labor (IZA), Bona, Discussion Paper N° 1664.

Vieira, J., Couto, J., and Tiago, M. (2006). "Regional differences in returns to education in portugal." *46th Congress of the European Regional Science Association University of Thessaly, Greece, August 30th-September 3rd*.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data - Second Edition*. The MIT Press, Cambridge, MA, Chapter 4, 5, 53–121.

A Additional tables of descriptive statistics

Table 11: Definition of the variables

Variable	Definition	Source
Lhwreal	Real hourly wage in logarithmic form (log Euro)	QP
Education	Years of Education and it were calculated by imputing the nominal number of years required to complete the level of schooling	QP
Occupation	Occupation level inside the firm is a dummy variable 1 = Executive and managerial 2 = Intermediate managerial and executive 3 = Low managerial 4 = Technicians highly specialized 5 = Sales, administrative and precision production 6 = Administrative support, and production 7 = Unskilled 8 = Apprentice 9 = Unknown	QP
Experience	Years of experience which is calculated by: $exper = age - years\,of\,education - 6$ if $(years\,of\,education + 6) \geq 16$ $exper = age - 16$ if $(years\,of\,education + 6) < 16$	QP
Tenure	Years of tenure (years of experience in the current firm) which is equal to the date of the questionnaire minus date of admission, converted to years	QP
Industry	Industry classification is a dummy variable for the sector of activity of the firm which the individual works and there are 18 classifications	QP
Attractiveness Index	Marktest Index which is considered a overall Index of Development of the county and it is computed as follows: $Attractiveness\,Index = \frac{PopulationIndex}{AreaIndex} \times 100 - 100$, where Attractiveness Index, Population and Area Index are expressed per thousand and positive values of the Attractiveness Index indicate that the county attracts population and negative values indicate the opposite	Sales Index 2013
Housing price	Ratio price of evaluation of accommodation per square meter by county between each county and Lisbon	INE
Unemployment rate	County unemployment rate	INE IEFP
Logpop	Log of the county population	Sales Index 2013
Log Aversize	Logarithm of average firm size in the county	QP

A.1 Additional tables of results

Table 12: Additional results of estimation for pooled OLS

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Log of real hourly wage				
Educ	0.0112***	0.0341***	0.0268***	0.0488***	0.0487***
	(0.0028)	(0.0011)	(0.0017)	(0.0005)	(0.0005)
Average educ by county	-0.0028				
	(0.0044)				
Educ*Educ county	0.0042***				
	(0.0003)				
Share of HQ workers by county		-0.0002			
		(0.0004)			
Educ*share HQ		0.0004***			
		(0.0000)			
Skill index by county			-0.0035		
			(0.0026)		
Educ*skill county			0.0025***		
			(0.0002)		
Log variation coefficient				0.0541*	
				(0.0291)	
Hirschman-Herfindahl index					-0.0000***
					(0.0000)
Female	-0.1636***	-0.1638***	-0.1631***	-0.1596***	-0.1598***
	(0.0035)	(0.0035)	(0.0035)	(0.0036)	(0.0036)
Exper	0.0205***	0.0205***	0.0204***	0.0209***	0.0208***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Exper squared	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0159***	0.0159***	0.0159***	0.0159***	0.0159***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Tenure squared	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	0.0461***	0.0461***	0.0461***	0.0464***	0.0459***
	(0.0027)	(0.0026)	(0.0027)	(0.0027)	(0.0027)
Log aversize	0.0200***	0.0202***	0.0207***	0.0250***	0.0227***
	(0.0047)	(0.0047)	(0.0046)	(0.0047)	(0.0048)
Log population	0.0052**	0.0063***	0.0056**	0.0039*	0.0034
	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Log attractiveness index	-0.0387***	-0.0439***	-0.0393***	-0.0120**	-0.0337***
	(0.0051)	(0.0053)	(0.0053)	(0.0050)	(0.0052)
Housing price (ratio)	0.0016***	0.0015***	0.0016***	0.0026***	0.0016***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0003)
Unemployment rate	-0.0028***	-0.0028***	-0.0031***	-0.0059***	-0.0047***
	(0.0009)	(0.0009)	(0.0009)	(0.0010)	(0.0009)
Occupation dummies	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes
Industry dummies	yes	yes	yes	yes	yes
Constant	1.4428***	1.4507***	1.4346***	1.1767***	1.5133***
	(0.0423)	(0.0428)	(0.0409)	(0.0387)	(0.0552)
Observations	14,142,939	14,142,939	14,142,553	14,142,939	14,142,939
R-squared	0.642	0.643	0.642	0.640	0.641
RMSE	0.340	0.340	0.340	0.341	0.341

Source: Computations of the author
Note: Robust standard errors in parentheses, clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Additional results of fixed-effects and random effects estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	**Dependent variable: Log of real hourly wage**											
	Individuals who did not changed the educational level						*Individuals who have change the educational level*					
	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE
Educ		0.0694***		0.0693***		0.0697***	0.0142***	0.0261***	0.0148***	0.0261***	0.0157***	0.0265***
		(0.0001)		(0.0001)		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Av. education	0.0069***	0.0236***					0.0247	0.0285***				
	(0.0001)	(0.0001)					(0.0003)	(0.0003)				
Share of HQ			0.0009***	0.0022***					0.0021***	0.0026***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Skill index					0.0051***	0.0134***					0.0109***	0.0158***
					(0.0001)	(0.0001)					(0.0001)	(0.0002)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	12,303,971	12,303,971	12,303,971	12,303,971	12,303,649	12,303,649	1,862,431	1,862,431	1,862,431	1,862,431	1,862,367	1,862,367
σ_α	0.5025	0.3057	0.5012	0.3057	0.5037	0.3058	0.3558	0.2543	0.3557	0.2542	0.3564	0.2543
σ_ϵ	0.1441	0.1434	0.1441	0.1434	0.1441	0.1434	0.1762	0.1740	0.1762	0.1740	0.1763	0.1740
ρ	0.9240	0.8195	0.9237	0.8195	0.9244	0.8197	0.8031	0.6811	0.8030	0.6809	0.8035	0.6810
χ^2 test		717,878		719,669		716,922		141,205		139,484		137,963
Degrees of freedom		30		31		31		31		31		31
Prob>chi2		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000
	Less qualified workers											
	Individuals who did not changed the educational level						*Individuals who have change the educational level*					
	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE
Educ		0.0454***		0.0455***		0.0458***	0.0090***	0.0179***	0.0092***	0.0179***	0.0097***	0.0184***
		(0.0001)		(0.0001)		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Av. education	0.0065***	0.0242***					0.0138***	0.0260***				
	(0.0001)	(0.0001)					(0.0003)	(0.0003)				
Share of HQ			0.0008***	0.0022***					0.0012***	0.0023***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Skill index					0.0042***	0.0134***					0.0062***	0.0136***
					(0.0001)	(0.0001)					(0.0002)	(0.0002)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	10,750,898	10,750,898	10,750,898	10,750,898	10,750,614	10,750,614	1,630,432	1,630,432	1,630,432	1,630,432	1,630,380	1,630,380
σ_α	0.3908	0.2754	0.3901	0.2754	0.3920	0.2754	0.3343	0.2482	0.3349	0.2482	0.3361	0.2483
σ_ϵ	0.1382	0.1375	0.1381	0.1375	0.1382	0.1375	0.1664	0.1656	0.1664	0.1656	0.1665	0.1656
ρ	0.8889	0.8004	0.8886	0.8005	0.8895	0.8005	0.8014	0.6921	0.8020	0.6920	0.8030	0.6922
χ^2 test		478,606		488,611		521,544		90,714		91,138		92,099
Degrees of freedom		30		31		30		31		31		31
Prob>chi2		0.0000		0.0000		0.0000		0.0000		0.0000		0.000

Continued on next page

Table 13 – continued from previous page

Most qualified workers												
	Individuals who did not changed the educational level						*Individuals who have change the educational level*					
	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE
Educ		0.1438***		0.1438***		0.1443***	0.0490***	0.0606***	0.0494***	0.0604***	0.0505***	0.0608***
		(0.0002)		(0.0002)		(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Av. education	0.0130***	0.0322***					0.0169***	0.0396***				
	(0.0003)	(0.0002)					(0.0007)	(0.0006)				
Share of HQ			0.0014***	0.0030***					0.0016***	0.0035***		
			(0.0000)	(0.0000)					(0.0001)	(0.0000)		
Skill index					0.0089***	0.0186***					0.0094***	0.0217***
					(0.0001)	(0.0001)					(0.0004)	(0.0003)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time dummies	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	4,289,883	4,289,883	4,289,883	4,289,883	4,289,845	4,289,845	580,663	580,663	580,663	580,663	580,651	580,651
σ_α	0.5372	0.3601	0.5368	0.3601	0.5370	0.3603	0.4266	0.3121	0.4270	0.3121	0.4283	0.3121
σ_ϵ	0.1651	0.1645	0.1651	0.1645	0.1651	0.1645	0.1810	0.1804	0.1810	0.1804	0.1810	0.1804
ρ	0.9137	0.8273	0.9136	0.8274	0.9137	0.8275	0.8475	0.7494	0.8477	0.7495	0.8485	0.7495
χ^2 test		257,285		254,381		250,231		37,486		37,509		37,883
Degrees of freedom		30		30		30		31		31		31
Prob>chi2		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000

Source: Computations of the author

Note: Robust standard errors in parentheses. For RE model, "χ^2" test is the Hausman test comparing the results with the FE model. *** p<0.01, ** p<0.05, * p<0.1