Human Capital, Geographical location and Policy Implications: The case of Romania

Jesús López-Rodríguez\textsuperscript{a,b}, Andres Faiña\textsuperscript{a,b}, Bolea Cosmin-Gabriel\textsuperscript{a}
\textsuperscript{a}Universidad de A Coruña
\textsuperscript{b}C+D Research Group

Abstract

This paper looks at the link between human capital and geographical location for the Romanian regions based on the theoretical model developed in Redding and Schott´s (2003) paper. Using 2006 data on the different educational attainment levels for the 42 Romanian regions, it identifies that the percentage of individuals with medium and high educational levels is affected positively by the regions´ market access. Doubling market access would increase the percentage of individuals with medium and high educational levels between 22-25%. Moreover the econometric results show that between 45% and 59% of the spatial variation in human capital levels is explained by the market access variable. Some policy implications to overcome the costs remoteness imposes on human capital accumulation in Romania are also drawn.

Key Words: Geographical location, Market Access, Human Capital, Romania

JEL Classification: R11, R12, R13, R14, F12, F23

* Corresponding author: Jesus Lopez-Rodriguez, Department of Economic Analysis and Business Administration, Faculty of Economics, Campus de Elviña, s/n. C.P.: 15071, A Coruña (Spain). Phone: +34 981 167050, ext: 2451, Fax: +34 981 167070, email: jelopez@udc.es
1. Introduction

Human capital can broadly be defined as “...the productive resources that focus on work resources, skills and knowledge" (OECD) or "human skills and capabilities generated by investments in education and health" (WHO). From these definitions it is clear that human capital must play an important role in the economic development of countries and regions. In fact, aggregate human capital at national or regional level has been a recurrent variable in economic growth models (Barro, 1991, 1997; Barro and Lee, 1994; Benhabib and Spiegel, 1994; Englander and Gurney, 1994; Hanushek and Kim, 1995; Islam, 1995). However, despite of the wide scholarly agreement of its impact on economic growth there is little consensus on the exact contributions of the different measures and indicators of human capital to economic development (Levine and Renelt, 1992, Rodriguez-Pose and Vilalta-Buffi, 2005). Another important issue related to human capital and economic development and far less studied is the role the economic geography of a country or a region plays with respect to this relationship. At this point the fairly new branch of the spatial economics known as New Economic Geography (NEG) (Krugman 1991, 1992) has emerged as a new theory which emphasizes the role second nature geography variables or economic geography variables play with respect to the spatial distribution of income and human capital across countries or regions as oppose to the role played by first nature geography variables (Hall and Jones, 1999). The emphasis of a large number of empirical studies in the NEG literature has been put on the effects economic geography have on either cross-country or cross-regional per capita income differences. This has been done by testing the well known theoretical proposition that arises in standard core-periphery NEG models which is refer to as the nominal wage equation ((Brakman et al. 2004, Breinlich, 2006, Hanson, 2005, Overman et. al., 2003, Redding and Venables, 2004, Lopez-Rodriguez and Faiña, 2007). However, recent theoretical developments within the NEG literature (Redding and Schott, 2003) has allowed to extend the empirical investigations to the analysis of the effects geographical location have on human capital accumulation.

1 By first nature geography we refer to the physical geography of a country (natural endowments, climate conditions, access to ports, airports, navigable rivers and so). Second nature geography refers to the economic geography, i.e. how far a country or region is from its consumer markets and from its input suppliers.
Redding and Schott’s (2003) pioneering paper extend a standard two-sector New Economic Geography model to demonstrate that being located on the economic periphery can reduce the return to skills, thereby reducing incentives for investment in human capital accumulation. To our Knowledge, the only empirical investigation of Redding and Schott’s (2003) paper was carried out by Lopez-Rodriguez et al. (2007) for a cross-regional setting of 205 NUTS2 regions in the European Union. Lopez-Rodriguez et al. (2007) paper provides evidence that in the case of the European Union (EU) there is a spatial educational attainment structure, i.e., educational levels are higher in regions located in the economic center of the EU than in the regions located in the economic periphery. This result jointly with those which relate the effects that economic geography has on cross-country and cross-regional per capita income can be considered as an additional prove of the penalty remoteness imposes for economic development and therefore for the convergence of countries and regions.

However, much more empirical studies on the relationship between human capital and location are needed. So far, there are no within country studies on the forces put at work in Redding and Schott’s (2003) paper.

This paper tries to fill in this gap by applying Redding and Schott’s (2003) framework to the regions in a national setting such as the case of Romania. The paper stresses, for the case of the 42 Romanian regions, the importance of geographical location in human capital accumulation, showing that the percentage of individuals with medium and high educational attainment levels depends positively on the region’s market access whereas the opposite occurs for low educational attainment levels. Moreover, the econometric results show that in Romania between 45% and 59% of the spatial variation in human capital levels is explained by the region’s market access.

The rest of the paper is structured as follows: section 2 contains the theoretical framework in which the relationship between human capital accumulation and geographical location is established. Section 3 presents the econometric approach and data. Section 4 discusses the econometric results on the link between educational attainment levels and remoteness. Finally section 5 presents the main conclusions and some policy implications.
2. Theoretical framework

The theoretical framework presented in this chapter is a short version of the Redding and Schott (2003) New Economic Geography model. The difference of our model with Redding and Schott’s (2003) model is in the modelling of the role played by intermediate goods. Contrary to Redding and Schott’s (2003) model we assume that the production of manufactured goods is carried out without using intermediates in the production of final output. The difference of this model with respect to standard two-sector New Economic Geography models such as Fujita et al. (1999) or Krugman (1991) is based on the introduction of endogenous human capital accumulation. To account for this new feature we consider a world in which we have R locations and each location have a mass of consumers $L_i$. We assume that consumers are endowed with one unit of labour which is offered inelastically with zero disutility and that consumers choose endogenously whether to invest or not in becoming skilled. In the decision of becoming skilled a worker has to compare the costs of education to acquire those skills with the future benefits of been skilled, which for the purposes of this paper can be summarized in the higher wages skilled workers perceive.

Therefore, the decision of an individual $z$ at location $i \in \{1, \ldots, R\}$ to become a skilled worker would be given by the wage differential between the two options, difference of wages of a qualified worker versus an unskilled one, and the costs associated with educate himself.

From a mathematical point of view, this condition can be expressed as follows:

$$w^s_i - w^u_i \geq \Omega_i(z)w^u_i$$  \hspace{1cm} (1)

Where $\Omega_i(z) = \frac{h_i}{a(z)}$ represents the costs of education in terms of units of unskilled labour to become skilled one. The cost function depends on two components, $h_i$, which is interpreted as an inverse measure of the extent of public provision of education and $a(z)$ which represents the ability of each individual. We understand that individuals’ ability is determined by human biology, so that the probability of finding individuals with a certain level of skills is the same for any location. From Equation (1) we obtain a value for $a$ so that if $a(z) \geq a^*(z^*)$, all individuals choose to become skilled.
\[ a_i^* = \frac{h_i}{(w_i^s/w_i^u - 1)} \] (2)

A worker with skill level noted as \(a_i^*\) is indifferent between becoming skilled or remain unskilled labour, so that this equation following Redding and Schott (2003) can be termed as the skill indifference condition (S).

In the same way as in standard models of New Economic Geography, this model assumes homothetic utility functions and the same preferences for all consumers, which are defined for the consumption of a homogeneous agricultural good and a set of differentiated manufactured goods. In this paper we skip the modeling of the demand side which is no necessary for the theoretical premises of our empirical investigation. Therefore, we focus on the agriculture and manufacturing equilibrium conditions (supply side), to characterize endogenously the relationship between geographical location and human capital accumulation.

The agricultural sector produces a homogeneous good under conditions of constant returns to scale. The production function can be given by the following expression:

\[ Y_i = \theta_i (S_i^Y)^{\phi} (L_i^Y)^{1-\phi}, \quad 0 < \phi < 1 \] (3)

\(Y_i\) represents the output of the agricultural sector. In this sector the output is produced using a \(\phi\) share of skilled workers and a \(1-\phi\) share of unskilled workers. \(\theta_i\) is a parameter representing the agricultural productivity in each location.

The manufacturing sector produces differentiated goods according to a technology which presents increasing returns to scale and where the production of each variety requires only primary factors of production (skilled and unskilled labour). The profit function of a typical firm at location \(i\) can be given by the following expression:

\[ \Pi_i = \sum_{j=1}^{R} \frac{P_{ij}^M x_{ij}}{T_{ij}^M} - (w_i^S)^a (w_i^U)^{1-a} c_i (F + x_i) \] (4)

Where \(P_{ij}^M\) is the price at location \(j\) of one unit produced at location \(i\), \(w_i^S\) is the wage of skilled workers with a share \((\alpha)\) in the total costs, \(w_i^U\) is the wage of unskilled...
workers with a share \((1 - \alpha)\) in the total costs, \(c_i\) is a marginal input specific to each location representing a technology index. \(F\) is a fixed cost of production and \(x_i = \sum_{j=1}^{R} x_{ij}\) is the total output produced by the company for all markets it serves.

Manufactured goods are traded between different locations incurring iceberg transportations costs, in other words a fraction of the good carried from location \(i\) to location \(j\) is melt in transit, so that for one unit to reach location \(j\) \(T_{ij} > 1\) units must be sent from \(i\) location.

Regarding to the producer’s equilibrium, the agricultural sector operates under a scheme of perfect competition which implies that price must be equal to the marginal costs of production:

\[
P_i^y = 1 = \frac{1}{\theta_i^y} (w_i^s)^{\phi} (w_i^u)^{1-\phi}
\]

As we choose the output of agricultural good as numeraire, we assign a price equal to 1 so that \(P_i^y = 1\) for all goods produced in different \(i\) locations.

Once we solve for the first order conditions of profit maximization, the expression in the manufacturing sector implies:

\[
(w_i^s)^{\alpha} (w_i^u)^{1-\alpha} = \xi c_i^{-1} (MA_i)^{\frac{1}{\sigma}}
\]

where \(\xi = \frac{\sigma}{1+\sigma}\) is a constant, \(c_i\) is the parameter that reflects differences in technology between locations, \(MA_i = \sum_{j=1}^{R} (T_{ij}^M)^{1-\sigma} E_j G_j^{\sigma-1}\) is the market access at location \(i\) and \(\sigma\) the elasticity of substitution between varieties of manufactured goods. The expression (11) is another way of conceiving the nominal wage equation from standard core-periphery New Economic Geography models

Combining equations (4) and (5) give us the equilibrium value for the wages of skilled and unskilled workers. Taking logarithms, totally differentiating expressions (4) and (5) and combining them with the skill indifference condition, allow us to obtain an
expression that relates geographical location with endogenous human capital investments.

\[ 0 = \phi \frac{dw_i^s}{w_i} + (1 - \phi) \frac{dw_i^u}{w_i} \quad (7) \]

\[ \alpha \frac{dw_i^s}{w_i} + (1 - \alpha) \frac{dw_i^u}{w_i} = \frac{1}{\sigma} \frac{dMA_i}{MA_i} \quad (8) \]

Considering equations (7) and (8) one can show that, if we make a shock so that the equilibrium value of market access decreases (\( MA_i \)), if the manufacturing sector is relatively skilled labour intense with respect to the agricultural sector, the new equilibrium is characterized by relatively lower wages of skilled workers. Therefore, this new equilibrium implies a higher critical level for the skill parameter (\( a_i^\ast \)) above which individuals prefer to invest in education and become skilled and thus we will have a lower supply of skilled workers. In this derivation we assume that the number of individuals with higher and higher levels of skills decreases as we move further away from those thresholds.

From the zero profit condition in the agriculture sector (Eq. 5) we can express the derivative of the wage of unskilled workers as follows:

\[ \frac{dw_i^u}{w_i^u} = -\frac{\phi}{(1 - \phi)} \frac{dw_i^s}{w_i^s} \]

If we now substitute this expression into the zero profit condition of the manufacturing sector we get the following expression. (Renamed \( 1 - \alpha = \beta \))

\[ (\alpha - \beta \phi) \frac{dw_i^s}{w_i^s} = \left[ \frac{1}{\sigma} \right]^\gamma \]

Knowing that: \( (\alpha - \beta \phi) > 0 \iff \frac{\alpha}{\beta} > \frac{\phi}{1 - \phi} \)

so

\[ \frac{dw_i^u}{w_i^u} > 0 \quad \frac{dw_i^s}{w_i^s} < 0 \quad \frac{d(\frac{w_i^s}{w_i^u})}{w_i^u} < 0 \]
This intuitive explanation is based on the fact that a decrease in market access modifies the initial equilibrium conditions in the manufacturing sector, which experience a decrease in size. This decrease in size, frees more skilled labour than the ones that are initially demand in the agricultural sector.

So to re-establish the equilibrium, the nominal wage paid to skilled workers must decrease and that nominal wages paid to unskilled workers must increase in relative terms. Therefore as the wages of skilled workers decrease this reduces the incentives to invest in becoming skilled.

3. Econometric Approach and Data

In this section we present the econometric approach we will use in the empirical estimations carried out in the next section of the paper. The theoretical propositions arising from the model can be estimating by running the following regression equation:

\[
\ln(\text{EA}_{i}) = \alpha_0 + \alpha_1 \ln(\text{MA}_{i}) + \varepsilon_i
\]  

(1)

\(\text{EA}_{i}\) represents the educational attainment level in region “i”, \(\text{MA}_{i}\) represents the market access for region i and \(\varepsilon_i\) represents the error term. Equation (1) allows us to check if there is a spatial educational attainment structure in Romania, i.e., namely whether there is a positive correlation between secondary and tertiary educational attainment levels and market access or alternatively if those regions which have a high market access index are also the regions with relatively high levels of education. We begin by examining how much of the variation in cross regional human capital can be explained when only including information on market access. This provides the basis for our baseline estimation where we assume that the error term is uncorrelated with the explanatory variables. Considering that this assumption can be violated and therefore the coefficient estimates be biased and inconsistent, we also present estimates using instrumental variables regression.

In order to control for the effects of outlying observations, we also estimate this alternative specification:

\[
\ln(\text{EA}_{i}) = \alpha_0 + \alpha_1 \ln(\text{MA}_{i}) + \sum_{n=1}^{N} \gamma_n X_{i,n} + \varepsilon_i
\]  

(2)
Where $X_i$ is a control variable and $\gamma_i$ is the correspondent coefficient.

To complement the estimations of different equations for different educational attainment levels, we also report the results of two alternative estimations based on transformations in the definition of the dependent variable. The first transformation of the dependent variable consists of ranking Romanian regions given the values 1 if low educational attainment is the highest share of educational attainment for a particular region and 2 if it is medium and high and then estimate an ordered probit model. The second transformation consists of estimating a single equation where the dependent variable is the average years of schooling in each region instead of educational attainments.

The dependent variable in the regression equation is the logarithm of educational attainment levels. We define two different types of educational attainment levels. In first place we consider the percentage of each Romanian region’s population that has attained secondary and tertiary education which will be labelled in the econometric estimations as $log$ Higher Education. In second place we define a new educational attainment level variable which takes in the percentage of each Romanian region’s population that has attained primary education which is labelled in the estimations as $log$ Lower Education. The former definition of the dependent variable, according to the model’s prediction, is a direct way to test for the validity of the forces put at work in the model whereas the latter definition of the dependent variable will constitute an indirect way to test model’s prediction. Both higher and lower educational attainment levels data are taken from the Romanian National Statistical Institute (INSSE) and refer to the year 2006.

The variable on the right hand side of expression (#1) is the regions’ market access. Taking into account that the market access of a region “i” is a distance-weighted sum of the volume of economic activity in the surrounding regions, we build a market access variable which takes as a proxy for the volume of economic activity the total gross domestic product in each region. For the calculation of the discount factor included in the market access variable, we use the distances measured in Kms between the capital cities of each Romanian region. Data on each region gross domestic product is taken from INSSE and refers to 2006 and the data for the distances between capital cities comes from the website www.travelworld.ro
For the calculation of the internal distance within each region, it is approximated by a function that is proportional to the square root of each region’s area. The expression used for calculation is \(0.66\sqrt{\frac{\text{Area}}{\pi}}\) where "Area" represents the size of the region expressed in km². This expression gives the average distance between two points on a circular location (see Crozet 2004, Head and Mayer, 2000, and Nitsch 2000) for a discussion of this measure of internal distance).

4. Empirical Analysis

Table 1 records 2006 data on the percentage of each Romanian region’s population that has attained primary education (labelled in the table as lower education) or secondary and tertiary education (labelled in the table as higher education). As it can be seen from the table, the educational attainment levels across Romanian regions vary greatly. The highest percentages of higher education are reach in the so called economic centers of Romania; Bucharest, Iasi, Timisoara, Cluj-Napoca, Constanta, Brasov and Craiova where also the country's main universities are located. The percentages figures on higher education in these regions are well above the Country’s average (8.55%) being Bucharest the region which ranks at the top (18.19%). On the other site, the Romanian regions located far from the above poles of growth in the so called Romanian economic periphery such as Piatra-Neamț Târgu Mureș, Tulcea, Satu Mare, Botosani, Vaslui, Olt, Teleorman have figures on higher education below the country’s average (6.97%).
<table>
<thead>
<tr>
<th>Region</th>
<th>Lower Education</th>
<th>Higher Education</th>
<th>Region</th>
<th>Lower Education</th>
<th>Higher Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacău</td>
<td>10,07</td>
<td>5,95</td>
<td>Drobeta T.</td>
<td>8,75</td>
<td>6,43</td>
</tr>
<tr>
<td>Botoşani</td>
<td>10,44</td>
<td>4,77</td>
<td>Saltina</td>
<td>9,28</td>
<td>4,88</td>
</tr>
<tr>
<td>Iaşi</td>
<td>9,91</td>
<td>12,74</td>
<td>Râmnicu Vâlcea</td>
<td>8,85</td>
<td>5,97</td>
</tr>
<tr>
<td>Piatra-Neamţ</td>
<td>9,50</td>
<td>1,97</td>
<td>Arad</td>
<td>8,48</td>
<td>7,87</td>
</tr>
<tr>
<td>Suceava</td>
<td>11,25</td>
<td>6,79</td>
<td>Reşiţa</td>
<td>8,68</td>
<td>8,37</td>
</tr>
<tr>
<td>Vaslui</td>
<td>10,66</td>
<td>6,12</td>
<td>Deva</td>
<td>8,59</td>
<td>7,19</td>
</tr>
<tr>
<td>Brăila</td>
<td>8,00</td>
<td>5,32</td>
<td>Timișoara</td>
<td>8,25</td>
<td>12,17</td>
</tr>
<tr>
<td>Buzău</td>
<td>8,62</td>
<td>4,67</td>
<td>Oraşea</td>
<td>9,12</td>
<td>9,43</td>
</tr>
<tr>
<td>Constanţa</td>
<td>8,55</td>
<td>9,97</td>
<td>Bistriţa-Năsaud</td>
<td>9,93</td>
<td>5,40</td>
</tr>
<tr>
<td>Galaţi</td>
<td>9,12</td>
<td>7,81</td>
<td>Cluj-Napoca</td>
<td>7,46</td>
<td>14,67</td>
</tr>
<tr>
<td>Tulcea</td>
<td>4,45</td>
<td>4,40</td>
<td>Baia Mare</td>
<td>8,80</td>
<td>6,49</td>
</tr>
<tr>
<td>Focşani</td>
<td>8,69</td>
<td>4,24</td>
<td>Satu Mare</td>
<td>9,68</td>
<td>5,32</td>
</tr>
<tr>
<td>Piteşti</td>
<td>4,58</td>
<td>7,98</td>
<td>Zalău</td>
<td>9,25</td>
<td>5,21</td>
</tr>
<tr>
<td>Călăraşii</td>
<td>9,09</td>
<td>4,46</td>
<td>Alba-Iulia</td>
<td>8,58</td>
<td>6,88</td>
</tr>
<tr>
<td>Târgovişte</td>
<td>9,32</td>
<td>6,19</td>
<td>Braşov</td>
<td>7,76</td>
<td>10,41</td>
</tr>
<tr>
<td>Giurgiu</td>
<td>8,92</td>
<td>2,80</td>
<td>Sfântu Gheorghe</td>
<td>8,97</td>
<td>5,59</td>
</tr>
<tr>
<td>Slobozia</td>
<td>9,22</td>
<td>5,25</td>
<td>Miercurea Ciuc</td>
<td>8,95</td>
<td>6,01</td>
</tr>
<tr>
<td>Ploieşti</td>
<td>7,99</td>
<td>6,10</td>
<td>Târgu Mureş</td>
<td>8,85</td>
<td>3,95</td>
</tr>
<tr>
<td>Alexandria</td>
<td>8,10</td>
<td>4,22</td>
<td>Sibiу</td>
<td>9,06</td>
<td>10,63</td>
</tr>
<tr>
<td>Craiova</td>
<td>8,46</td>
<td>8,91</td>
<td>Butea</td>
<td>0,90</td>
<td>3,02</td>
</tr>
<tr>
<td>Târgu Jiu</td>
<td>9,95</td>
<td>8,03</td>
<td>Bucureşti</td>
<td>5,91</td>
<td>18,19</td>
</tr>
</tbody>
</table>

Moreover, these figures on the spatial distribution of educational attainment levels across Romanian regions show a well established core-periphery gradient, a pattern that is commonly observed when we refer to the analysis of the spatial distribution of incomes (poor regions predominantly located in the so called “economic periphery” whereas rich ones are located in the so called “economic center”). Figure 1 illustrates this fact by plotting the percentage of population with higher education (in logs) in 2006 against distance from one of the Romanian economic centers (Timișoara).
Before presenting the results of the econometric estimations carried out with 2006 data for the Romanian regions, we proceed presenting a couple of graphs which relate different levels of regional educational attainment in Romania and the corresponding regional market access. Figure 2 plots the percentage of individuals with secondary and tertiary education in each Romanian region (log Higher Education) against each Romanian region market access. As it can be seen in the graph the pairs of values (Higher Education, Market Access) are distributed along a positive slope trend line indicating that higher market access regions have higher levels of secondary and tertiary education. The relationship higher education-market access is robust and not due to the influence of a few regions. Therefore, figure 2 corroborates, at least graphically, the theoretical predictions of the model.
Finally, an indirect way (graphically) to check for the validity of the theoretical predictions of the model is to plot primary educational attainment levels against market access and see how the set of points (primary education, market access) are distributed. This has been done in figure 3. The graph clearly shows that the set of points are distributed along a negative slope trend line, meaning that those regions with higher levels of market access have lower percentages of individuals with primary education or alternatively as the regions remoteness increases the incentives to become skilled diminish and therefore we found lower levels of individuals with higher education.

Source: Own elaboration using data from INSSE
The previous descriptive analysis characterizes the relationship between different classifications of the educational attainment levels in Romania and market access. In this section we extend the analysis with a regression model. Taking into account our theoretical framework OLS and Instrumental Variables regressions of secondary and tertiary educational attainment levels for the year 2006 are conducted on the Romanian regions’ market access. Market access has been computed by using gross domestic product as the proxy of the volume of economic activity for each Romanian region and labelled in the table as MAGDP06.
Table 2 presents the results of estimating equation (1) on the sample of 42 regions in Romania for the year 2006. In Column 1 we regress Log Higher Education on market access for the set of 42 Romanian regions. The results of the OLS estimation show that the coefficient of market access has the expected sign and is statistically significant at the 1% level. The results also show that doubling regions’ market access would increase secondary and tertiary education attainment levels by 25%. The null hypothesis that the coefficient on market access is equal to zero is easily rejected at conventional significance levels using a standard F-test, and the model explains over 59% of the cross-regional variation in secondary and tertiary educational levels.

In column 4 we summarize the results of regressing the percentage of population with primary education (labelled as Log Lower Education in the table) against market access. The results of the OLS estimation indicate that an increase in regional market access is
negatively correlated with the percentage of population who has primary education. This result constitutes an indirect way of checking the theoretical predictions of the model.

A potential shortcoming of the previous analysis is the one referring to the endogeneity of the market access measure, i.e., good market access can be correlated with other determinants of the level of educational attainment of the Romanian regions and therefore cause inconsistent and biased estimates. To avoid problems of endogeneity between human capital levels and regional market access, the paper presents instrumental variables estimates. IV estimation is based on the existence of a set of instruments that are strongly correlated with the original endogenous variables but asymptotically uncorrelated with the error term. Furthermore, they should also be variables that are not driven by an unobservable third variable the authors suspect might be jointly affecting market access and human capital levels. Once these instruments are identified, they are used to build a proxy for the explanatory endogenous variables which consists of their predicted values in a regression on both the instruments and the exogenous variables. However, it is difficult to find such instruments because most socioeconomic variables are endogenous as well. In this paper we propose to use mainly accessibility variables as instruments, since they are highly correlated with our market access variable but also non contemporary correlated with the errors. Following Breinlich (2006), in this paper we instrument market access with distance from Timisoara and with the region size. The first instrument capture market access advantages of regions close to the geographic centre of Romania. The second instrument captures the advantage of large regional markets in the composition of domestic market access.

Columns 2 and 5 present the results for the corresponding instrumental variables estimation. Instruments are highly statistically significant and have the expected signs in the first stage. Distance to Timisoara and regions size explains 62% of regional market access. Since the instruments represent quite a distinct source of information and are uncorrelated, we can trust them to be reliable instruments. In the second-stage estimation we again find positive and highly statistically significant effects of market access on educational attainment levels although its effects are lower than in the OLS estimations. The market access coefficients change from 0.25 to 0.22 in the regression
of log higher education against market access (column 2) and from -0.15 to -0.17 in the regression of log lower education against market access (column 5).

For comparison purposes, column 3 reports the result of regressing log higher education against distances from Timisoara instead of using market access. The result provides evidence of the negative correlation between secondary and tertiary educational attainment levels and regions distance from Timisoara.

The estimation of two different equations *log Lower Education* and *Log Higher Education* is based on the fact that the coefficient estimates are significantly different for the two equations. In order to check this fact we run this alternative regression:

$$\ln(EA_{ij}) = \alpha_0 + \alpha_1 \ln(MA_{ij}) + \alpha_2 D_{ij} + \varepsilon_{ij}$$

(3)

Where $i = 1,2,...,42$ represents the 42 Romanian regions of our sample, $j = \{0,1\}$ stands for the level of educational attainment, being 0 if educational attainment is defined as lower education and 1 if educational attainment is defined as higher education, so $EA_{i,0}$ is the proportion of population in region 1 who has primary educational levels and $EA_{i,1}$ is the proportion of population in region 1 who has secondary and tertiary educational levels. $MAGDP06_{ij} = MAGDP06$, for all $j = \{0,1\}$ is the market access of region $i = 1,2,...,42$ and $D_{ij} = \{0,1\}$ is a variable that takes the value 0 if $j = \{1\}$ and 1 if $j = \{0\}$, $\varepsilon_{ij}$ stands for the error term.

In this alternative specification our main parameter of interest is $\alpha_2$ such that if $\alpha_2$ is statistically different from cero, we can reject that the estimated coefficient $\alpha_i$ is equal for the different equations and thus it confirms our approach to the problem. The results reported in column 6 of table 2 shows that $\alpha_2$ is significantly different from cero, thus justifying the estimation of two different equations for the different levels of educational attainments.
However, the models given in table 2 are marked by outlying observations. The outlying regions do not correspond with the spatial educational attainment structure determined by the majority of the observations. Outliers will seriously affect the coefficient estimates, if they are influential leverage points, i.e. outlying observations with regard to our market access measure. We identify outliers as those observations for which Cook's distance is greater than 1. In order to control for the effects of the identified outlying observations, dummy variables for the outliers are introduced. The most significant outliers are the Romanian capital, Bucharest and the regions of Târgu Mureș, Buftea and Târgu Jiu.

The first column of table 3 reports results of regressing log lower education on log market access for the 42 Romanian regions after including dummies for the outlying observations. The estimated coefficient on market access is negative and statistically significant at the 1% level. The second column of Table 3 shows the results of the estimations of log higher education against log market access. The result is robust and the market access coefficient is again significant at the 1% level. The third column of table 3 indicates that market access retains a significant positive relationship with higher education even in the presence of indicators thought to be important in cross regional development in Romania. The indicators, all referring to 2006 and available from INSSE, we use consist of the expenditure in R&D expressed as percentage of regional Gross Domestic Product, the share of ethnic minorities in the population of each region and the average gross monthly earnings. Including these variables in column (3) reduces the magnitude of the market access coefficient from 0.30 to 0.13 although it remains statistically significant at conventional critical values. Among the controls, only the expenditure in R&D is statistically significant.
To complement our estimations columns 4 and 5 of table 3 summarize the results of two alternative estimations based on transformations in the definition of the dependent variable. In column 4 we transform Romanian regional educational attainment levels into average years of schooling and then we estimate a single equation using average years of schooling as our dependent variable. This synthetic indicator for human capital levels has been used in many empirical studies see (Benhabid and Spiegel 1994, Temple 1999, Krueger and Lindahl 1999 and De la Fuente and Domenech 2001). To do the transformation of educational levels into average years of education we use information

### Table 3: Market Access, Regional Dummies, Educational Levels and Average Years of Education Romania (2006)

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>Log Lower Education</th>
<th>log Higher Education</th>
<th>log Higher Education</th>
<th>Average Years Education</th>
<th>Educational Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regress.</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.53**</td>
<td>0.92**</td>
<td>-2.24</td>
<td>6.01**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.16)</td>
<td>(3.65)</td>
<td>(0.35)</td>
<td></td>
</tr>
<tr>
<td>MAGDP06</td>
<td>-0.15**</td>
<td>0.30**</td>
<td>0.13**</td>
<td>0.60**</td>
<td>1.82**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.10)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>R&amp;D Expenditure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minorities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regional</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>IV</td>
<td>IV</td>
<td>OLS</td>
<td>OLS</td>
<td>Ord. Probit</td>
</tr>
<tr>
<td>Inst. variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First stage R2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.59</td>
<td>0.61</td>
<td>0.68</td>
<td>0.51</td>
<td>0.71</td>
</tr>
<tr>
<td>J-Statistic</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Prob (F-statistic)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.obs.</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>

**Note:** Table displays coefficients and Huber-White heteroscedasticity robust standard errors in parenthesis, ** indicates coefficient significant at 0.01 level, “First stage” R2 is the R2 from regressing market access on the instruments set, Instruments: Distance to Timisoara and region size.

Source: Own Elaboration
of the Romanian school system provided by the Ministry of Education, Research and Innovation. Romanian school system consists of the pre-university education system and the university education system. The pre-university education is broken down into 4 levels (preschool, primary, secondary level 1, secondary level 2). Primary education covers 4 courses and students are enrolled at the age of 6 and finish at the age of 10. Secondary education is divided into two additional levels (level 1 and level 2) each of them of 4 years length; level 1 from 10 years old to 14 and level 2 from 14 to 18: Finally the higher education includes vocational training, usually three years, from 18 to 21 and university education which in Romania is on average 4 years length.

The results of the regressions show that the coefficient on market access is positive and statistically significant at the usual critical values, showing that an increase in a regions’ market access increases the average years of education of its population. Column 5 summarize the results of estimating an ordered probit model where the dependent variable was transformed into a binary variable given to it the values 1 or 2 according to the relative importance of the proportion of population who has low or medium or high educational levels. Therefore a region that has the highest proportion of population with low education is ranked 1, if the highest proportion is secondary and tertiary education is ranked 2. In ordered probit models, the sign of the coefficient shows the direction of the change in the probability of falling in the endpoint rankings, in our case (Educational attainment level 1, lower education, or level 2, higher education) when market access changes. Probability of Educational Attainment level 1 changes in the opposite direction of the sign of the estimated coefficient and probability of educational attainment level 2 changes in the same direction. The coefficient reported in column 5 of table 3 is positive showing that the probability of having higher educational levels is higher in regions with high market access. The estimated coefficient is statistically significant at the conventional critical values2.

Therefore the results reported in columns 4 and 5 can be taken as additional proofs that geographic location matters for determining educational levels across Romanian regions.

---

2The statistic reported in ordered probit models to check the significance of the estimated coefficient is z-statistic instead of t-statistic from OLS.
5. Conclusions and Some Policy Implications

In this paper we use 2006 data on Romanian regional educational attainment levels to look at the link between human capital accumulation and geographical location. The theoretical framework of the paper, based on Redding and Schott (2003), presents a model which is an extension of the standard two-sector (agriculture and manufacturing) Fujita et al. (1999) economic geography model in which unskilled individuals are allowed to endogenously choose whether to invest in education. The main theoretical result of the model proves that relatively peripheral locations will experience a lower skill premium and therefore this reduces their incentives to educate their workers. Consistent with the predictions of the model, our empirical findings emphasize the importance of economic geography in explaining the spatial structure of the Romanian regional human capital levels. The results of the bivariate regression of secondary and tertiary educational attainment levels against market access (regression of log higher education on log market access) show that the coefficient estimates of market access are positive and statistically significant. This result shows that high market access regions are endowed with higher levels of individuals with secondary and tertiary education which is in line with the theoretical predictions of the model. In particular the results show that if we double the market access of a region, the percentage of individuals with higher education would increase between 22-25%. Moreover around 59% of the spatial variation in higher education is explained by the regions market access. The results of the bivariate regression prove to be robust to the inclusion of dummies and to the inclusion of other indicators important in cross-regional development in Romania such as regional expenses in R&D, the presence of ethnic minorities in the region’s population and the gross average monthly earnings. The results of the extended regressions (including dummies and other regional indicators) affect the coefficient estimates of market access reducing its magnitude from 0.30 to 0.13 although it remains statistically significant at conventional critical values. We also check indirectly the model’s prediction by regressing the percentage of individuals with primary education against market access (Log lower education on log market access). The results of the estimations show a statistically significant negative coefficient for market access which means that as the regions market access increases the percentage of individuals with low educational attainment levels decreases. This backs indirectly the results of the direct estimates. Finally we complement our estimations with two alternative estimations
based on transformations in the definition of the dependent variable. In the first case we use average years of education as our dependent variable and in the second case the dependent variable was transformed into a binary variable given to it the values 1 or 2 according to the relative importance of the proportion of population who has low educational levels or medium or high educational levels. The results of these alternative regressions back again the main results found in the paper.

The results of our paper have also important implications in policy terms for Romania. Based on the fact that remoteness hampers human capital accumulation which is considered a key engine to fuel economic growth and therefore to accelerate the development of countries and regions, an obvious policy implication is that remote locations in Romania need to get closer to the centers of economic activity. Though locations cannot move, is it possible to reduce the costs of remoteness?. Perhaps most important in this regard will be the policy actions to reduce transport costs directly via improvements in infrastructure (e.g. roads, ports, etc.) which in the case of Romania are still lagging behind.

The recent accession of Romania to the European Union will mean that in the years to come it will receive big amounts of funding via Structural Funds and Cohesion funds. An important policy priority therefore should be to channel part of these funds to tackle the infrastructural problems Romania is facing.

However, the Romanian accession to the European Union imposes also some challenges. With free movement of goods, people and capital, the risks of a "brain drain" of highly qualified people to other member states with better salaries is a fact that has been taken place ever since the Romanian access to the European Union. Moreover, other important issues that may hamper Romanian human capital accumulation in the short and medium term are among others the negative demographic trends characterized by low birth rates and high mortality rates, the overall health situation, the dropout rates which are relatively high, the low level of adult participation in lifelong learning, the large proportion of the population engaged in agriculture, particularly subsistence agriculture, the high unemployment above all long-term youth unemployment and the matching problems between the educational offer and what the job market really needs. Therefore, a clear strategy to overcome these problems establishing the right priorities with respect to the Romanian human resources is also needed. In this respect again an important role should be played by the European Union structural funds. As is stated in
the current programming period (2007-2013), the Romanian strategy on human resources development wants to eliminate or reduce these weaknesses. Another important challenge refers to the management of the European funds. Good managerial practices must be set up in order for the European funds to deliver the expected results and to pursue the goals established at the 2005 March summit of the European Council “Europe must renew the basis of economic competitiveness and to increase potential growth and productivity, strengthen social cohesion, placing greater emphasis on knowledge, innovation and optimization of human capital”.
6. References


European Observatory on Demography and the Social Situation: http://ec.europa.eu/employment_social/social_situation/sso_en.htm


Ministerul Economiei și Comerțului, Programul Operațional Sectorial 2006 “Creșterea Competitivității Economice” (www.minind.ro)


OECD: Organization for Economic Co-operation and Development

Organizația Mondială a Sănătății (WHO): “Health status overview for countries of Central and Eastern Europe that are candidates for accession to the European Union”, 2002.


