AFC ASSOCIATION FRANÇAISE DE CLIOMÉTRIE

WORKING PAPERS

Nr. 8, 2016

Remoteness equals backwardness?
Human capital and market access in the European regions: insights from the long run

Claude Diebolt and Ralph Hippe

Remoteness equals backwardness?

Human capital and market access in the

European regions: insights from the long run

Claude Diebolt[1] and Ralph Hippe[2]

Abstract

In a recent contribution, Redding and Schott (2003) add human capital to a two sector NEG model, highlighting that remoteness represents a penalty that gives disincentives to invest in human capital. But is this hypothesis consistent with long-term evidence? We test the persistence of this effect at the regional level in an historical setting. The results show that market access has a significant positive influence on human capital in OLS, Tobit and IV regression models. Thus, the paper confirms the 'penalty of remoteness' hypothesis for Europe in the long run.

Keywords: Human Capital, New Economic Geography, Regional Development, Market Access.

JEL Codes: I21, N33, N93, R11.

[1] BETA/CNRS, Université de Strasbourg; cdiebolt@unistra.fr, claude.diebolt@hu-berlin.de.
[2] BETA/CNRS, Université de Strasbourg; hippe@unistra.fr.

1. Introduction

Human capital is generally perceived to be a key factor for today's knowledge-driven economies. This is particularly true for Europe and the European Union. For this reason, the Council of the European Union highlights that "[e]ducation and training have made a substantial contribution towards achieving the long-term goals of the Lisbon strategy for growth and jobs" (Council of the European Union 2009, C 119/2). Still, the EU is facing important challenges in its regional policy. Although the EU has aimed to decrease economic and social inequalities over the last decades, there still remain important differences between and within countries. The current economic crisis has further widened previous convergence tendencies. Similarly, education is not equally distributed in space. Thus, how can one explain these differences?

One possible explanation advanced by theory and in particular by models from New Economic Geography (NEG) is that consumer markets play an important role in the distribution of economic development. These models have already been tested empirically for the last decades (e.g., Breinlich 2006, Faíña and López-Rodríguez 2006, Head and Mayer 2011) and have confirmed the predictions provided by these rather recently developed NEG models. A particular case including human capital formation is presented by Redding and Schott (2003). The authors develop a theoretical NEG model showing that remoteness from large consumer markets gives disincentives to individuals to increase their human capital. For this reason, this 'penalty of remoteness' explains worldwide inequalities in human capital accumulation. Subsequent empirical studies have also confirmed the predictions of the model for the European regions for the last couple of years (e.g., López-Rodríguez *et al.* 2007).

Nevertheless, (to our knowledge) there has not yet been any empirical evidence for the long-term evolution of market access and human capital at the EU or European

regional level. This empirical evidence, however, appears particularly important to understand the changes that have shaped today's European regions in the long run. This may considerably enlarge the recent analyses for the short term which may be only part of a much larger long-term process.

For these reasons, this paper explores for the first time the importance of market access for the spatial distribution of human capital in the European regions in the past. We combine and adapt several databases to create a new unique dataset. More specifically, we use two different human capital indicators at different points in time to test the robustness of our analysis. First, we employ regional numeracy estimates for 1850. The age heaping method enables us to estimate numeracy (e.g., A'Hearn *et al.* 2009, Hippe and Baten 2012, Hippe 2012a). Second, we use literacy as an alternative human capital proxy. Literacy is certainly the most employed indicator for human capital in the past. Therefore, we can check the overall numeracy results by using regional literacy outcomes in 1930. Both indicators also allow to better model the proposed theory than alternative historical education variables. In addition, as has been proposed by the recent literature, we exploit data on the distribution and size of cities in Europe to model historical market access.

The results show that market access has a significant negative influence in OLS, Tobit and IV regressions. In the latter case we use distance to Luxembourg and area size of European countries as instrumental variables. We also control for outlying regions in several specifications. In sum, the 'penalty of remoteness' hypothesis theoretically advanced by Redding and Schott (2003) is confirmed by our historical data. This result implies that the 'penalty of remoteness' is not a current trend but has existed for long time spans, the present being only a very special case of a larger phenomenon.

The paper is structured as follows. First, we consider the literature on human capital formation in the European regions in the past and the main contributions of New

Economic Geography. Then, we briefly present the underlying theoretical NEG framework which has been originally proposed by Redding and Schott (2003). Subsequently, the data and the econometric specifications are discussed. In the fourth section we show the results. The final section concludes.

2. Related literature

2.1 Regional human capital formation in Europe, today and in the past

Human capital formation in the European regions has attracted the attention from many researchers (e.g., Badinger and Tondl 2003, Breinlich 2006, Faíña and López-Rodríguez 2006, Sterlacchini 2008). For example, Rodríguez-Pose and Tselios (2011) use Exploratory Spatial Data Analysis to test the spatial distribution of educational attainment in western Europe between 1995 and 2010. They find that educational attainment is strongly correlated with inequality and that regions tend to cluster in space. Proximity plays an important role for educational attainment even today. Moreover, there are noticeable differences between the north and the south of western Europe and the urban and rural communities.

However, as the authors state, "[t]he geography of education, especially at [the] subnational level, is a huge black box" (Rodríguez-Pose and Tselios 2011, p. 358). If this is still true for today, one can imagine how the situation is for the past. New evidence on the regional distribution of human capital in Europe in the past has recently been provided by Hippe and Baten (2012). They use the age heaping method to calculate numeracy estimates, i.e., whether individuals are able to count or calculate (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010). They show that regional numeracy values steadily improved almost everywhere in Europe during the 19th century. Leaders in numeracy were countries in Scandinavia, in central Europe and the United Kingdom. Many of the regions in these

parts of Europe had already very high numeracy values at the beginning of the 19th century. In contrast, Southern and Eastern Europe importantly lagged behind. They needed many more decades to attain similar levels, in part even until the beginning of the first or the second half of the 20th century. In addition, regional differences in numeracy were quite striking in most of these countries. For example, a core-periphery pattern characterised Spain. The regions in the north of Madrid had the highest numeracy levels while those at the southern periphery (Andalusia) and in the north-western periphery (Galicia) followed with a large gap. In contrast, a north-south gap is visible in Italy. The lowest numeracy values were calculated for regions in the Balkans and the Caucasus. Still, most countries and regions were able to improve the numeracy levels of their population during the 19th century. In consequence, regional inequalities in numeracy diminished over that period.

Numeracy is one possibility to measure human capital in the past but there are also other proxies. However, one has to note that no other measure is available at such a scale to that time period. Nevertheless, one can focus on individual countries to check the validity of the numeracy data. In fact, the most important tendencies in numeracy can also be detected when employing other indicators. For instance, Cinnirella and Hornung (2011) use data on Prussian counties in the 19th century. Taking a look at the data one can see that the counties of Poznan province had the lowest enrolment rates (6-14 year olds), confirming the lower numeracy levels in the study by Hippe and Baten (2012).

Another country example is provided by Felice's (2012) recent study on the regions of Italy during the 19th and the 20th century. In contrast to the above-mentioned studies, he uses a specifically constructed human capital measure which takes into account both enrolment rates and literacy. He shows that regional differences in human capital peaked around 1871 but diminished during the next decades. Northern regions had a

distinct lead to other regions, followed by central and ultimately southern and island regions.

Furthermore, Núñez' (1992) Spanish literacy data underline a core-periphery structure similar to the one highlighted by numeracy. Finally, the correspondence of regional numeracy and literacy data has been put emphasised by Hippe (2012b), using data for a number of European countries around the beginning of the 19th century and adding more recent data for current developing countries in Latin America, Asia and Africa during the second half of the 20th century.

2.2 NEG and market access in Europe

New Economic Geography (NEG) has become an important field in economics over the last years. NEG models enable to understand why economic activity and individuals cluster in space (e.g., Krugman 1991a). In other words, they allow to clarify the reasons for the existence of urban agglomerations, e.g., Tokyo and Mexico City, and areas with concentrated activity, such as the Manufacturing Belt in the United States and the Blue Banana in Europe. In fact, concentration is the most evident characteristic of economic geography (Krugman 1991b). Accordingly, the regional distribution of GDP per capita in Europe is quite unequal.

The spectacular growth of urban agglomerations, particularly in developing countries, further shows that economic geography is an important factor for the distribution of the population in the past, today and probably in the future. Given these facts, it is not astonishing that policy makers are faced with the question of how to deal with these inequalities. Economic geography in general and NEG in particular has gained attention due to the process of European integration and its consequences for regional inequalities (Fujita *et al.* 1999). In this area, the concept of market access or market potential plays a very important role in many NEG models (e.g., Crozet 2004, Redding and Venables 2004,

Hanson 2005, Niebuhr 2006). Having a good access to large markets is deemed to be a fundamental economic advantage of a region. The notion of market access, however, is older than NEG. Harris (1954) initially proposed the concept. Later applications include Clark *et al.* (1969) and, in particular, Keeble *et al.* (1982) (see also Niebuhr 2006). The latter authors show that market potential is lowest in periphery regions. The highest market potential, in contrast, was found in north-western Europe, including West Germany and the Benelux countries. However, these studies lacked solid theoretical foundations. These were only later provided by NEG, so that newer studies are able to test the implications of these theoretical models. Initially, these were country studies (e.g., Roos 2001, Brakman 2004, Mion 2004, Ottaviano and Pinelli 2006) which generally emphasise the importance of market access. More recently, new studies also take a European approach (e.g., Head and Mayer 2006, Niebuhr 2006) and generally confirm the hypotheses set up by NEG.

Market access may also have effects on the accumulation of human capital. Human capital is clearly an important economic factor which may enable higher growth rates and lead to convergence or divergence processes. However, the incentive for individuals to invest in their human capital and their geographic location are not independent. In particular, higher market access may encourage human capital accumulation. This hypothesis has been validated by a range of publications focussing on the worldwide national level (e.g., Redding and Schott 2003) and later on the European regional level. In particular, the latter include the work by López-Rodríguez and co-authors (e.g., López-Rodríguez *et al.* 2005, Faíña and López-Rodríguez 2006, López-Rodríguez *et al.* 2007). The results of these papers clearly indicate that human capital levels (as approximated by educational attainment levels) decrease when moving from NUTS 2 regions with high market access to those with low market access in the year 2000. Therefore, market access is a crucial factor influencing human capital formation in the

present time. However, has this always been the case? Is it a more general pattern that has persisted until the present time? This paper contributes to answer this question by analysing econometrically the importance of market access in the long run.

3. Theoretical model

The proposed NEG model has originally been developed by Fujita *et al.* (1999). This model has two sectors, i.e., agriculture and manufacturing. However, the model does not take into account human capital accumulation. This factor has only been added by Redding and Schott (2003). Their model focuses on the interaction between human capital and input-output linkages, taking account of transport costs and assuming increasing returns to scale. One of their main results is that countries that are remotely located from main markets have to face higher trade costs and a decrease in the skill premium than other countries if one assumes that manufactures are relatively more skill intensive than agricultural goods. In this way, the effect of a remote location has the same consequences as a reduction in the relative price level of manufactures. Due to the assumption that the required skills in the manufacturing sector are higher than in agriculture, skilled workers face a fall in their relative wages. Thus, the incentive for an unskilled worker to invest in human capital and become skilled is decreased.

Because the main contribution of this paper is empirical, we only briefly present some foundations and results of Redding and Schott's (2003) model.[3] We adapt the model to the context of this paper by explicitly considering regions (instead of countries as in the original model). First, we consider the preferences and the endowments that have to be modelled. Accordingly, Europe is constituted by $i \in \{1, \dots, R\}$ regions. Every region is

[3] We follow López-Rodríguez *et al.* (2007) and limit ourselves to the supply side of the model. For a complete presentation, see e.g., López-Rodríguez *et al.* (2005).

characterised by an endowment of L_i consumers. Every consumer has one single unit of labour. The supply of this unit of labour is inelastic, i.e., there is no disutility. Consumer preferences are identical for all L_i. Consumption is restricted to two types of goods: first, the production of the agricultural sector is limited to one homogenous good. Second, the manufacturing sector produces a range of differentiated manufactures. The preferences follow a standard utility function in Cobb-Douglas form.

Let us now define the production technologies involved in the two sectors. In the first sector, the produced agricultural good is homogeneous. Production is set within the framework of perfect competition and is characterised by constant returns to scale,

$$Y_i = \theta_i^Y (S_i^Y)^{\phi} (L_i^Y)^{1-\phi}, \qquad 0 \prec \phi \prec 1 \tag{1}$$

where Y_i is the agricultural sector's output, L_i^Y is the quantity of unskilled workers in the sector, S_i^Y is the quantity of skilled workers and θ_i^Y is an index of agricultural productivity. In the second sector, the production of the differentiated manufactured goods is characterised by increasing returns to scale and uses a combination of the two types of labour (skilled, unskilled) and of the intermediate inputs of manufactured goods.

In the next step, we introduce endogenous investment in human capital into the model. It is assumed that a conversion from an unskilled to a skilled worker is possible. Denoting an individual as *z*, this conversion incurs a fixed cost of education $\Omega_i(z)$ units in terms of unskilled labour. The underlying idea is that real resources are consumed to become skilled which results in the fact that the education cost is a proportion of the wage of unskilled labour. Moreover, the quantity of unskilled labour that is needed to become skilled is dependent on two factors. In particular, $\Omega_i(z) = \frac{h_i}{a(z)}$, where h_i denotes the overall environment provided by institutions and government policies that have repercussions on the education cost and $a(z)$ denotes the individual's personal ability. This

ability is subject to human biology. Thus, an individual z will only take the decision to invest in human capital if

$$w_i^S - w_i^U \geq \frac{h_i}{a(z)} w_i^U, \tag{2}$$

i.e., if education costs are lower than (or equal to) the difference between the wages of a skilled (w_i^S) and an unskilled (w_i^U) worker. The equation defines an implicit critical value for a above which all individuals choose to invest in human capital. This value a_i^* giving the supply of skills in equilibrium is

$$a_i^* = \frac{h_i}{\left({w_i^S}/{w_i^U} - 1 \right)}. \tag{3}$$

An individual having the ability a_i^* does neither prefer to become skilled nor to remain unskilled but is indifferent to both options. Therefore, this equation is the 'skill indifference condition'. Only if an individual has an ability above a_i^* he will choose to get further education.

After defining the producer equilibrium and profit maximisation, we can obtain the zero-profit condition, i.e. the wage equation. It defines the maximum level of wages that receive both categories of workers that firms in *i* are able to pay:

$$(w_i^S)^\alpha (w_i^U)^{1-\alpha} = \xi \frac{1}{c_i} (MA_i)^{\frac{1}{\sigma}}, \tag{4}$$

where ξ is a composite constant, c_i stands for an inverse indicator of technological efficiency, α is the input share of skilled workers wages, σ denotes the elasticities of substitution and MA_i is the market access of *i*.

Combining the zero-profit conditions for agriculture and manufacturing gives the wages that are paid to workers with and without skills in equilibrium. If we join the skill indifference condition, we obtain the equilibrium relationship that exists between the

geographical location of a region and endogenous investments in human capital. Thus, we have,

$$0 = \phi \frac{dw_i^S}{w_i^S} + (1 - \phi) \frac{dw_i^U}{w_i^U} \tag{5}$$

$$\alpha \frac{dw_i^S}{w_i^S} + (1 - \alpha) \frac{dw_i^U}{w_i^U} = \frac{1}{\sigma} \frac{dMA_i}{MA_i} \tag{6}$$

These equations show that if the equilibrium MA_i decreases, if the manufacturing sector is assumed to be skill intensive with regard to the agricultural sector and if the region is incompletely specialised, then the equilibrium moves to a new equilibrium with lower skilled wages but higher unskilled wages. This implies that the critical ability level a_i^* increases. This change induces a lower supply of skilled and a higher supply of unskilled labour.

More specifically, the decrease of MA_i has led to a smaller size of the skill intensive manufacturing sector. The reduction in size means that there are now more skilled workers in the market than there is demand for them in agriculture. Therefore, the wages of skilled workers decrease whereby their relative wages in comparison to the ones of unskilled workers fall. In this way, remoteness leads to smaller incentives to invest in human capital. This means that the model predicts a positive relationship between market access and human capital investment.

While this basic two sector model has so far been used to explain the current economic geography of Europe, it appears even more appropriate for the past. Clearly, whether to switch from agriculture to manufacturing is not a policy question for most European countries anymore. The service sector has become much more relevant, both in terms of GDP and employment, than agriculture in many European countries. In contrast, the simple structure of the model may even more closely mirror the development during the European industrialisation process. Most European countries only began to

industrialise during the 19th or even the 20th century. For example, Broadberry (2009)'s data show that agriculture had still a share of 50 % in agricultural employment in West Europe in 1870. The share increases to 57 % in South Europe and even 70 % in East Europe. Without taking major assumptions, it is evident that these shares would be even higher for 1850, illustrating the crucial stake of agriculture in Europe at that time. Ongoing industrialisation in West Europe increases the share of industry while decreasing agricultural employment to 32 % in 1929. Nevertheless, a third of overall employment is still a relevant share for agriculture, particularly if compared to only 5 % of employment in West Europe in 1992 (and certainly less today). However, industrialisation occurred later and slower in other, peripheral parts of Europe. Therefore, the same share only dropped to 53 % in South Europe and to 66 % in East Europe in 1929. Therefore, the theoretical model's two sector model and the associated switch from unskilled to skilled workers appears related to the historical period under study (i.e., 1850 and 1930). The higher market access in the core industrialising European countries would imply that skilled workers are rarer in the periphery. Can we find this theoretical result also in the data?

4. Data and methodology

We test the theoretical model by the use of different datasets. In particular, we use regional numeracy and literacy as our human capital proxies. First, we employ numeracy as a proxy for regional human capital in Europe in 1850. Numeracy is derived from the age heaping method. Age heaping as a method for calculating basic human capital values has been established by the recent literature (e.g., A'Hearn *et al.* 2009, Crayen and Baten 2010, Hippe and Baten 2012, Hippe 2012b). In particular, we use the ABCC Index to measure numerical abilities. In fact, it measures the share of individuals that are able to calculate. More specifically, historical census data, and in part even data for today's LDCs, show a

clear pattern of rounding. Many people were not able to calculate their age. Therefore, they guessed their age to fulfil the census requirements set up by the state. Given that human biology serves as a first aid for calculations (e.g., five fingers on one hand, ten fingers in total), they rounded their ages on 0 and 5 (see also Harper 2008). It has been shown that this rough proxy is well correlated with other standard human capital proxies such as literacy (A'Hearn *et al.* 2009, Hippe 2012b) and primary school enrolment (Crayen and Baten 2010). The underlying formula of the ABCC Index is

$$ABCC_{it} = 125 - 125 \times \left(\sum_{j=5}^{14} n_{5j,it} \Big/ \sum_{j=23}^{72} n_{j,it} \right), \tag{7}$$

where *i* denotes a region, *j* the number of years, *n* the number of individuals and *t* the time period. Its formula illustrates that one calculates the share of age observations ending on '0' and '5' in relation to all observations. One takes into account all ages between 23 and 72, the standard in the numeracy literature. The ABCC with its limits of 0 and 100 is comparable to other share indexes, in particular literacy.

The human capital data have been taken from the new and large database provided by Hippe and Baten (2012). These data are based on original historical census data. The advantage of this measurement method is that it always takes into account the entire population and not, as other historical proxies of human capital (e.g., signature rates), only parts of it. For this reason, it is representative for the whole population and is not prone to biases that naturally reside in more partial indicators. In this way, we are able to measure the regional distribution of basic numeracy from Portugal to Russia. In total, there are 299 regions in our dataset (see Table 1 for descriptive statistics).[4]

[4] Given the variable "Distance to Luxembourg", we have excluded Luxembourg in all our regressions and do not list it here.

Second, literacy is our alternative human capital proxy and available for 1930. Literacy data are not available for a range of European regions for 1850, which is why the ABCC Index is the more suitable indicator for that period. However, literacy became a standard human capital indicator during the second half of the 19th century and was used in many countries throughout the first half of the 20th century. Many public debates focused on the eradication of illiteracy in a number of European countries. Still, this aim was sometimes not achieved until the second half of the 20th century, so that it is a valuable indicator for 1930. Therefore, it can be used as the representative human capital proxy for 1930. Literacy is defined as

$$Literacy_{it} = \sum_{j=10}^{N} rw_{j,it} \Big/ \sum_{j=10}^{N} n_{j,it}, \tag{8}$$

where rw denotes the ability to read and write and N is the total number of years. In other words, literacy is the share of individuals (10+ years) who are able to read and write in a region at a given point in time. Data stem originally from Kirk (1946) and have been adapted for the purposes of this paper.

Both numeracy and literacy have the advantage that they are share variables. In fact, the proposed NEG model divides individuals into skilled and unskilled workers. A defined level of numeracy is nothing more than the share of numerate to innumerate, a level of literacy is the share of literate to illiterate individuals. We can take the simple but straight forward assumption that an unskilled worker is innumerate (in 1850) or illiterate (in 1930). Similarly, it is reasonable to assume that an innumerate person can decide to become numerate and an illiterate one to become literate. The endogenous investment assumption in the model, from unskilled to skilled workers, can thus be illustrated by our indicators. These parallels show the correspondence between our empirical specification and the underlying theoretical model. Therefore, our binary human capital or skill proxies

follow more closely the theory than other potential education variables, such as years of schooling.[5]

Moreover, the data on urbanisation are provided by two different sources. For 1850, we use the data provided by Bairoch *et al.* (1988). It is, alongside with a similar database by De Vries (1984), the standard database on urbanisation in the long run. In fact, the data trace back the cities of Europe until the year 800, starting from 1850. For a general geographical illustration of the data for 1850 see Figure 1. London is by far the largest European city, followed by Paris. Cities are quite dense in most of Europe, except Scandinavia and Eastern Europe, where Russia's capital, St. Petersburg, is the most important city. Cities are included if they fulfil a minimum threshold of population size between 800 and 1800. This threshold is 5000 inhabitants. In total, there are 2201 cities in our database. We excluded two observations because they were geographical outliers so that we have used the remaining 2199 cities for our calculations.[6]

Because the Bairoch *et al.* database does not cover later points in time, we had to use another database for our literacy regressions. In fact, European wide data for literacy are only available around 1900 onwards but the earliest data on cities (or, in this case, agglomerations) after 1850 are only available for 1950. Therefore, we use literacy data from 1930 and take as the best approximation of market access in 1930 data on European agglomerations in 1950.[7] These agglomeration data have been assembled by Moriconi-

---

[5] In addition, these variables are not available at the different points in time for most of Europe. Similarly, the distinction between skilled and unskilled workers is less clear cut for educational attainment levels. Even if it was available (which is not the case), given the low extent of higher or even secondary education in Europe in the 19th and large parts of the 20th century, it would not be a representative skill proxy either.

[6] These outliers are Ponte Delgada which is on the Azores Islands and far off the European continent. Moreover, we excluded Oral which is not located in the limits of today's definition of Europe.

[7] We are very well aware of the fact that World War II affected important portions of regional populations which may have a biasing effect on our estimates. However, authors such as Martí-Henneberg (2005) show that population concentrations are highly correlated at the regional level between 1870 and 2000 which suggests that data from 1950 are still a good approximation for 1930.

Ebrard (1994) and are compatible to the Bairoch *et al.* database.[8] It is a worldwide database with a threshold of 10000 inhabitants in 1990, so that the entire database includes up to 26000 worldwide agglomerations. A graphical illustration of these data for Europe shows their general resemblance to those in 1850 (see Figure 2). While London is still the most populous European city, there have been increases in population elsewhere. For example, Russia's new capital, Moscow, is now significantly larger than St. Petersburg and other capitals and agglomerations such as Athens show an increased importance in the European urban landscape. However, the overall picture that we get from 1950 is quite similar to the 1850 data.

Furthermore, market access has been proxied by population potential in the recent literature (e.g., López-Rodríguez *et al.* 2005).[9] Population potential also appears to be the best available proxy in historical European applications.[10] It is a standard way of representing changes in the pattern in which cities are distributed in space. It allows to identify the relative location of a city within a greater network of other cities. Two factors are essential in the evaluation process: first, the size of the population of cities. Second, the distance of a city to the other regions in the network. In practice, one adds to the population size of a city the population sizes of the other cities, each time divided by their distances to the original city. This is done for every city in the data. In this way, a potential value is assigned to each city. To be more precise, the mathematical formula in correspondence to López-Rodríguez *et al.* (2005) is

---

[8] Moriconi-Ebrard's (1994) database includes agglomeration data from 1950 to 1990.

[9] Clearly, it would be preferable to use an even closer theory-based measure, including regional price and interregional trade flow data. Yet, as López-Rodríguez *et al.* (2007) already emphasise, this measure is not available for today. Without surprise, it is not available for the past either, so that we have to rely on our alternative but fairly good proxy estimates.

[10] Other economic measures, such as regional GDP data, are not yet available for an important part of European regions, in particular in Eastern Europe.

$$MA_i = Po_i + \frac{Po_i}{D_{i,1}} + \cdots + \frac{Po_n}{D_{i,n}} = Po_i + \sum_{j \neq i, j=1}^{N} \frac{Po_j}{D_{i,j}}, \tag{9}$$

where MA_i stands for the market potential at i, Po_i is the population of i and $D_{i,j}$ is the distance that exists between i and j, each i and j representing individual nodes.

For the econometric specification of the relationship between investment in human capital and market access, we first test a standard OLS regression model as used by the literature. The basic framework is the following:

$$\ln(HC_i) = \beta_0 + \beta_1 \ln(MA_i) + \varepsilon_i, \tag{10}$$

where *HC* is the respective human capital indicator (i.e., numeracy or literacy; in logarithmic terms), *MA* is the market access (in logarithmic terms), *i* is a region and ε are the unexplained residuals. The basic OLS framework is later complemented by Tobit and instrumental variable regressions.

In addition, note that 'region' stands for a NUTS region in our case. NUTS is the official Nomenclature for Territorial Units of Statistics which has been developed by the European Union. It comprises all countries of the EU, EFTA and Candidate Countries of the EU. For countries outside this area, e.g., Russia, we used the current administrative division. This allows us to make our data comparable to current data and other research. Given the fact that market access and distance involves point data (cities and the central point of each region, respectively), the NUTS level can be attributed without any further difficulties. The case is different for human capital data which were available only for the historical regions. In this case, we developed the correspondence of these historical regions to current regions as best as possible. Because we have often more detailed data than needed for this study (e.g., the *départements* in France, the *provincias* in Spain or the *Bezirks-Hauptmannschaften* in Cisleithania (i.e., the Austrian part of Austria-Hungary before its disestablishment), the possible biases are importantly reduced because we can

easily aggregate our data from province or county level. As a standard, we use NUTS 2 as the basic unit of analysis which is also the standard unit in most other contributions in our area.[11] In this way, we are able to create a unique dataset for the European regions in 1850 and 1930.

5. Results

We first present our results for numeracy in 1850 and subsequently check their overall robustness with the literacy data in 1930. The calculated population potential values for 1850 are illustrated in Figure 3. In the following we refer to countries and regions in their current boundaries. It is apparent that the highest population potentials are found in the areas of Paris, London and Manchester and the wider locus up to Belgium and the western parts of Germany. In the current literature, this area is often called Golden Triangle and has also been identified by market access studies such as López-Rodríguez *et al.* (2007) for 2000. The similarity of our historical market access estimates to their current estimates further shows the validity of our approach and of potentially existing long-term spatial configurations. Given the size of the aforementioned cities, in particular of Paris and London, the shape of the triangle is not surprising because these were the two most populated cities of Europe in 1850. Still, the figure highlights that they were not isolated from other population hubs but were the centre of a greater accumulation of population in western Europe. This can be explained by the long-term geographic change of economic importance from northern Italy to this area, as has also been postulated by Braudel (1979). This is also in line with the concept of the existence of a “blue banana” which has been put

[11] An exception is Greater London, where we had to use the NUTS 1 level due to unavailability of more disaggregated data.

forward by Brunet (2002), a concentration of population and economic active from northern Italy over the course of the Rhine River until the UK and even Ireland.

In general, the more one distances oneself from the centre in western Europe, the lower are the potential population values. Going farther away from the centre, the highest estimated values are located in the regions of the UK, France, Germany, Switzerland, Italy, parts of Austria and Spain. Polish regions are already in the next level. Nevertheless, there are some outliers to the overall rule. Large cities create their own high local population potential which explains the different shading in the areas of e.g., Madrid, Hamburg, Berlin, Prague, Vienna, St. Petersburg and Moscow.

In the next step, we investigate the relationship between market access and human capital. To this end, we plot market access against the ABCC (Figure 4). Unfortunately, the ABCC has already achieved its maximum level of 100 in several countries. This is why there are a number of regions that are limited by the upper bound. Nevertheless, there is a clear relationship between market access and the ABCC. Outlying regions are in particular Greater London (UKI) at the very right and Albania at the bottom of the figure. Thus, the scatterplot allows us to identify the most relevant outliers that we should incorporate in our econometric models.

To test this relationship econometrically, we perform different regression models. As OLS is the most basic and standard estimation method, we begin with OLS regressions. Subsequently, we will also test alternative models that incorporate issues concerning the scale of the dependent variable (i.e., Tobit models) and regarding endogeneity (i.e., instrumental variable models).

The results of the baseline OLS regressions are shown in Table 2. Market access has a highly significant positive effect on numeracy at the 1 % level (column 1).[12] A 1 % increase in market access increases numeracy by 0.17 %, a sizeable effect. To compare our results for market access with distance to Luxembourg as proposed in the literature, we also computed this distance (in natural logarithm) and show the results in column 2 (and also in the subsequent steps). Distance to Luxembourg is negatively significantly correlated to the ABCC at the 1 % level. To avoid biases, we also include dummies for the most important outlying regions in our data as López-Rodríguez *et al.* (2007). As identified above, these regions are London (with by far the highest market potential) and Albania (with by far the lowest ABCC) (column 3). Both London and Albania are negatively significant at the 1 % level. The case for Albania is clear due to its very low ABCC. London is at the top of the market potential scale. Still, it has not achieved the maximum attainable level of numeracy, which one may expect from its substantial market potential. Therefore, its coefficient is negative but not very large. The inclusion of these outliers improves considerably the explanatory power of the model, as the increase in the R^2 underlines (from 0.27 to 0.44 in the case of the market access specifications, see column 1 and 3).

However, we have seen in the scatter plot that there are a number of regions that have already achieved the upper bound of 100 ABCC points in 1850. This given upper limit may bias our results because some of these regions would have had higher numeracy values if the limit was not existing. For this reason, we take this fact explicitly into account by running the same regressions with the Tobit model. The Tobit model incorporates the problem of upper or lower bounds in its estimations. The lower bound is not important in

[12] Note that we have opted for the presentation of the results with the logarithmic form of the ABCC. We have also done all regressions without this transformation and obtained the same results (only the value of the coefficients changed which is a logical consequence of the transformation).

our case, but the upper limit is. Thus, in total, there are 41 regions which are right-censored by the model.[13] The results when using the alternative Tobit model are shown in columns 5 to 8. The coefficient of market access increases slightly from 0.17 to 0.18 in the new specification while remaining significant at the 1 % level (columns 5 and 7). The coefficients of distance to Luxembourg (column 6) and of the dummies also increase (column 7). Comparing the distance to Luxembourg model including dummies (column 8) to its equivalent OLS specification (column 3) reveals that the inclusion of the upper bound has turned the London dummy insignificant and its coefficient dropped to zero. Apart from this expected result, there are not any large differences to our OLS estimates. In consequence, the Tobit model confirms the robustness of our former results.

Nevertheless, it is still possible that our results are biased by endogeneity. In fact, one can imagine that market access is correlated with alternative variables that may have a significant influence on numeracy. Thus, to be able to identify whether there is causality between market access and numeracy, we also perform instrumental variable regressions. In the given case, an instrumental variable has to be a determinant of market access but also has to be exogenous to numeracy. Moreover, the variable should not be prone to influences of another underlying variable which may drive its values and affect both market access and numeracy.

Thus, in line with Redding and Venables (2004), Breinlich (2006) and López-Rodríguez *et al.* (2007), we take the distance from Luxembourg as our first instrumental variable. This variable captures the advantages conferred by being close to the centre of Europe. Second, as proposed by the same authors, we use the (area) "size of a region's home country" (López-Rodríguez *et al.* 2007, p. 223), capturing the advantages that are

[13] Because we use the logarithmic form of the ABCC here, the upper limit (corresponding to 100) is approximately 4.6052.

created by big national markets for the market access of a region.[14] The use of a similar strategy as previous authors also enables us to put their results for today into a larger historical context, which is the aim of this paper. The results of our IV models are shown in column 9 and 10. The IV estimates for (logarithmic) market access are once again highly significant at the 1 % level. The signs of all coefficients do not change and the level of the coefficients is close to the one in our other specifications. In other words, the coefficient of (logarithmic) market access was 0.17 in the OLS, 0.18 in the Tobit and now increased slightly to 0.19 in our IV models. The coefficients of the dummy variables remain within the limits of the OLS and Tobit models and thus barely change. In sum, the IV results confirm once more the importance of market access for numeracy.

However, one may wonder if our results are robust to the use of other human capital variables and other time periods in the past. Therefore, our alternative indicator for human capital in the past is literacy in 1930. For this reason, our results would need to be confirmed by the use of this alternative indicator. However, given the use of another dependent variable (i.e., another human capital proxy), the consideration of a later time period (i.e., 80 years later than our numeracy estimates), and another dataset for the calculation of market access (although compatible with the dataset for 1850), we clearly would not expect to obtain the same results, including the same level of coefficients. In particular, the scatterplot has shown that literacy rates are much more dispersed than numeracy rates. For this reason, we expect higher coefficients in our 1930 regressions. Nevertheless, we expect to come to the same broad conclusions using this alternative specification.

---

[14] Borders and countries in ca. 1850 are considered. Because we are interested in the domestic market and trade advantages, we consider Germany as being constituted by those countries that had joined the *Zollverein* (German Customs Union). Data on country sizes (in geographical square miles) come from Annuaire Statistique et Historique Belge (1857).

To achieve a maximum of comparability with our earlier results, we take the same approach as for numeracy in 1850. First, we find that the results of the population potential calculations appear to be quite similar around 1950 (see Figure 5). The ‘core’ of population potential is still located within the Golden Triangle, i.e. the industrial areas of England, Paris, Belgium and western Germany. The Iberian Peninsula, Scandinavia and eastern Europe makes still up the periphery. Some differences emerge, however. For example, there appeared to be only two significant city centres with a high population potential in eastern Europe, that is St. Petersburg and Moscow. Now, these two cities are joined by Donetsk. In contrast, the cities in Spain and Portugal are not so relevant outliers anymore. The higher fertility rates and increasing urbanisation in eastern Europe over the previous century may explain these changes. Still, the overall pattern is quite robust to these rather minor changes. In addition, it appears to be still more closely related to current market access estimations by López-Rodríguez *et al.* (2007). Similar to their data for the year 2000, Romanian Bucharest is now a positive outlier (accompanied by Bulgarian Sofia). Major population potential levels are now extending until Polish Wroclaw and Italian Milan, again quite similar to the recent data for 2000. These findings give additional validity and show the robustness of our estimations and may indicate of the long-term nature of regional market access levels.

Next, plotting market access and literacy shows that their positive correlation is also quite clear (see Figure 6). Note that there are no literacy data for several developed countries in 1930 such as the Scandinavian countries, Germany or the UK. Kirk (1946) estimates that these countries had literacy rates between 95 and 100. In the following, we exclude the regions from these countries (as has been done in Figure 6). Alternatively, we can also take the hypothesis that these regions had a literacy rate of 100. In any case, there are no apparent outliers. The relationship between literacy and market access is even closer

than for numeracy. Therefore, we do not need to include additional dummies as in the previous numeracy analysis.

The corresponding regression results are shown in Table 3. This time, we propose two different specifications. First, we exclude the developed countries without any official literacy data (column 1 to 3). Column 1 shows that log market access is again positive and highly significant at the 1 % level. That is, a 1 % increase in log market access increases literacy by 0.56 %. This is substantially more than the 0.17 % that we obtained for numeracy in 1850. As noted above, this higher level corresponds to our expectations. All coefficients are higher than in 1850 because literacy rates are more dispersed than numeracy rates. A similar reasoning applies to the higher negative and significant coefficient for log distance to Luxembourg (column 2). As we have excluded all estimated literacy data for developed countries, the remaining countries do not reach the upper bound of 100 % literacy. For this reason, we do not need to perform Tobit regressions. Even if we perform them, we get the same results (not shown). Therefore, we proceed with the IV estimation, using the same strategy as in our numeracy regressions (column 3). The coefficient of log market access remains highly significant at the 1% level and largely stable, increasing only slightly from 0.56 to 0.57. This is the same tendency we have already observed in our numeracy sample.

Second, we include the developed countries with their estimated literacy rates (column 4 to 8). Most of these countries have estimated literacy rates of 95-100 %.[15] Thus, we assume that these countries had literacy rates of 100 %. As now a number of countries have reached the upper limit, we perform Tobit analyses in addition to OLS models.[16] Although we prefer to exclude countries that have the same estimated literacy rate for each

[15] See Kirk (1946) for more information.

[16] These countries are Denmark, Germany, Ireland, Netherlands, Norway, Sweden, Switzerland, United Kingdom and parts of Austria.

region, as we did in our previous specification, this alternative strategy may allow us to show the effects of including all of Europe. We start again with OLS models (column 1 and 2). While the significance levels are identical, the coefficient of log market access (column 4) decreases from previously 0.56 to 0.51 due to the higher number of observations (with a high level of literacy). However, log distance to Luxembourg (column 5) remains unchanged. Moving to Tobit models, the coefficient of log market access increases from 0.51 (column 4) to 0.77 (column 6) by about 50 %. However, the same appears for log distance to Luxembourg (column 7). In consequence, taking account of the upper limit increases the coefficient even more than when excluding the countries with estimated high literacy rates (see the initial models in column 1 to 3) because we add a high number of regions that have achieved the upper limit (without the estimated literacy rates we have 201 observations (column 1), now we have 327 (column 4)). Finally, we perform the IV regression which brings us to the same level as in the initial OLS models without the estimated regions. In other words, we obtain a positive and highly significant coefficient of 0.57, the same level as in our previous IV estimation in column 4, although we have included a higher number of (estimated) observations. While the coefficients are higher, as expected, in our literacy regressions for 1930, the significance of the results remains robust. In consequence, market access is a highly significant determinant of human capital in every model.

All in all, we thus find a core-periphery pattern in Europe similarly to the literature that analyses the EU today. Market access has a significant influence on human capital, confirming the ‘penalty of remoteness’ hypothesis. Moreover, because we are referring to the rather distant past with our data, the current regional distribution of human capital and economic development appears to be rather stable in the longer run.

This gives important implications for regional policy. Remote regions need to obtain better access to the main European markets. For this reason, it appears to be essential to advance improvements in infrastructure and focus even more on investments in human capital.

6. Conclusions

This paper has analysed the importance of market access to explain the spatial distribution of human capital levels in the European regions in the long run. The central focus of the paper is whether remoteness was connected to backwardness in the past, as has been postulated by Redding and Schott (2003) and tested by e.g., López-Rodríguez *et al.* (2007) for the European regions in the present.

In particular, we construct a new combined dataset using two different indicators of human capital, numeracy and literacy, to check the robustness of our results. More specifically, we employ, first, the age heaping method in order to approximate numeracy values for 1850. Second, literacy is a standard human capital proxy in Europe for the end of the 19th and the first half of the 20th century. Therefore, we use literacy in 1930 as our alternative specification. These two binary indicators also partly mirror the assumptions of the underlying theoretical model. Moreover, data on European cities have been used to proxy for market access. In this direction, the standard concept of population potential has been employed to generate average market access estimations for the European regions.

The results show that market access is highest in the regions of the Golden Triangle, i.e., England, northern France, Belgium and western Germany. In general, the farther one moves away from this centre, the lower is the level of market access. Therefore, we find a core-periphery structure also in the past.

Moreover, OLS, Tobit and IV regressions of market access on numeracy highlight that numeracy is significantly higher in regions with higher market access. We also control for outlying regions which improves the explanatory power of the model. Thus, after the literature has in particular used educational attainment for the current period, our numeracy and literacy estimates show that the 'penalty of remoteness' hypothesis is not only valid for today but that its importance can be traced back even to the middle of the 19th century.

This underlines once more that this penalty has existed for a long time in Europe. Thus, it may continue to exist also in the future if not the right policy decisions are taken. Improved infrastructure and greater incentives for investment in human capital appear to be very important in this context.

References

A'Hearn, B., Crayen, D. and J. Baten (2009). Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital, *Journal of Economic History*, 68 (3): 783-808

Annuaire Statistique et Historique Belge (1857). *Quatrième année*, Brussels/Leipzig/Bonn: Aug. Schnée et Cie./A.D. Marcus

Badinger, H. and G. Tondl (2003). Trade, human capital and innovation: the engines of European regional growth in the 1990s, in: Fingleton, B. (ed.). European Regional Growth, Heidelberg, New York: Springer, 215-239.

Bairoch, P., Batou, J. and P. Chèvre (1988). *La population des villes européennes : 800-1850 : banque de données et analyse sommaire des résultats*, Geneva: Droz

Brakman, S., Garretsen, H. and M. Schramm (2004). The spatial distribution of wages: estimating the Helpman–Hanson model for Germany, *Journal of Regional Science*, 44: 437-466

Braudel, F. (1979). *Civilisation materielle, économic el capitalisme, XVe-XVIIIe siècle. Le temps du monde*, Paris: Armand Colin

Breinlich, H. (2006). The spatial income structure in the European Union: what role for economic geography?, *Journal of Economic Geography*, 6: 593-617

Broadberry, S. (2009). Agriculture and structural change: lessons from the UK experience in an international context, in: Lains, P. and V. Pinilla (eds.). Agriculture and economic development in Europe since 1870, London and New York: Routledge.

Brunet, R. (2002). Lignes de force de l'espace européen, *Mappe Monde*, 66,2002.2

Cinnirella, F. and E. Hornung (2011). Landownership Concentration and the Expansion of Education, EHES Working Papers in Economic History, No. 10, October 2011.

Clark, C., Wilson, F. and J. Bradley (1969). Industrial location and economic potential in Western Europe, *Regional Studies*, 3: 197-212

Crayen, D. and J. Baten (2010). Global Trends in Numeracy 1820-1949 and its Implications for Long-Run Growth, *Explorations in Economic History* 47: 82-99

Crozet, M. (2004). Do migrants follow market potentials? An estimation of a new economic geography model, *Journal of Economic Geography*, 4: 439-458

De Vries, J. (1984). *European Urbanization, 1500-1800*, London: Taylor and Francis

Eurostat (2009). *Eurostat regional yearbook 2009*, Luxembourg: Publications Office of the European Union, online, last accessed on 15 August 2014, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-HA-09-001-01/EN/KS-HA-09-001-01-EN.PDF

Faíña, A. and J. López-Rodríguez (2006). Market access and human capital accumulation: the European Union case, *Applied Economics Letters*, 13: 563-567

Fujita, M., Krugman, P. and A. Venables (1999). *The Spatial Economy: Cities, Regions and International Trade*, Cambridge: MIT Press

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration, *Journal of International Economics*, 67: 1-24

Harper, D. A. (2008). A bioeconomic study of numeracy and economic calculation, *Journal of Bioeconomics*, 10: 101-126.

Harris, C. (1954). The market as a factor in the localization of industry in the United States, *Annals of the Association* of *American Geographers*, 44: 315-348

Head, K. and T. Mayer (2004). The Empirics of Agglomeration and Trade, in: Henderson, J. V. and J.-F. Thisse (eds.). Handbook of regional and urban economics, Vol. 4, 2609-2669, Oxford: Elsevier

Head, K. and T. Mayer (2006). Regional wage and employment responses to market potential in the EU, *Regional Science and Urban Economics*, 36: 573-594

Hippe, R. (2012a). Spatial Clustering of Human Capital the European Regions, *Economies et sociétés*, 46 (7): 1077-1104

Hippe, R. (2012b). How to measure human capital? The relationship between numeracy and literacy, *Economies et Sociétés, AF*, 45 (8): 1527-1554

Hippe, R. and J. Baten (2012). Regional Inequality in Human Capital Formation in Europe, 1790 – 1880, *Scandinavian Economic History Review*, 60 (3): 254-289

Keeble, D., Owens P. L. and C. Thompson (1982). Regional accessibility and economic potential in the European Community, *Regional Studies*, 16: 419-432

Kirk, D. (1946). *Europe's Population in the Interwar Years*, Princeton: Princeton University Press

Krugman, P. (1991a). Increasing returns and economic geography, *Journal of Political Economy*, 99 (3): 483-499

Krugman, P. (1991b). *Geography and Trade*, Cambridge: MIT Press

López-Rodríguez, J., Faíña, J. A. and J. López-Rodríguez (2005). New Economic Geography And Educational Attainment Levels In The European Union, *International Business & Economics Research Journal*, 4 (8): 63-74

López-Rodríguez, J., Faíña, J. A. and J. López-Rodríguez (2007). Human Capital Accumulation and Geography: Empirical Evidence from the European Union, *Regional Studies*, 41 (2): 217-234

Martí-Henneberg, J. (2005). Empirical Evidence of Regional Population Concentration in Europe, 1870-2000, *Population, Space and Place*, 11: 269-281

Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy, *Journal of Urban Economics*, 56: 97-118

Moriconi-Ebrard, F. (1994). *Geopolis, pour comparer les villes du monde*, Economica-Anthropos, Paris: Colle. Villes

Niebuhr, A. (2006). Market access and regional disparities, *Annals of Regional Science*, 40: 313-334

Ottaviano, G. I. P. and D. Pinelli (2006). Market potential and productivity: Evidence from Finnish regions, *Regional Science and Urban Economics*, 36: 636-657

Redding, S. and A. Venables (2004). Economic geography and international inequality, *Journal of International Economics*, 62: 53-82

Redding, S. and P. Schott (2003). Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich?, *Journal of Development Economics*, 72 (2): 515-41

Rodríguez-Pose, A. and V. Tselios (2011). Mapping the European regional educational distribution, *European Urban and Regional Studies*, 18 (4): 358-374

Roos, M. (2001). Wages and market potential in Germany, *Jahrbuch für Regionalwissenschaft*, 21: 171-195

Sterlacchini, A. (2008). R&D, higher education and regional growth: Uneven linkages among European regions, *Research Policy*, 37 (6): 1096-1107

Table 1 Descriptive statistics for ABCC and market access, ca. 1850

Variable	obs.	mean	sd	min	max
ABCC	299	90.81	12.27	26.38	100.00
Market access	299	4465.12	2200.10	995.13	17165.47
Distance	299	14.19	11.36	0.61	51.50

Table 2 Market access and ABCC, ca. 1850

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable						ln(ABCC)				
ln(MA)	0.17***		0.17***		0.18***		0.18***		0.19***	0.19***
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	(0.000)
Dummy London			-0.18***	0.03***			-0.21***	0.00		-0.21***
			(0.000)	(0.000)			(0.000)	(0.679)		(0.000)
Dummy Albania			-1.20***	-1.18***			-1.21***	-1.19***		-1.19***
			(0.000)	(0.000)			(0.000)	(0.000)		(0.000)
ln(Dist. to Lux.)		-0.10***		-0.09***		-0.12***		-0.11***		
		(0.000)		(0.000)		(0.000)		(0.000)		
Constant	3.12***	4.72***	3.12***	4.71***	3.01***	4.78***	3.02***	4.78***	2.93***	2.94***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Estimation	OLS	OLS	OLS	OLS	Tobit	Tobit	Tobit	Tobit	IV	IV
Observations	299	299	299	299	299	299	299	299	299	299
Prob. (F-Statistic)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.27	0.25	0.44	0.42					0.27	0.45

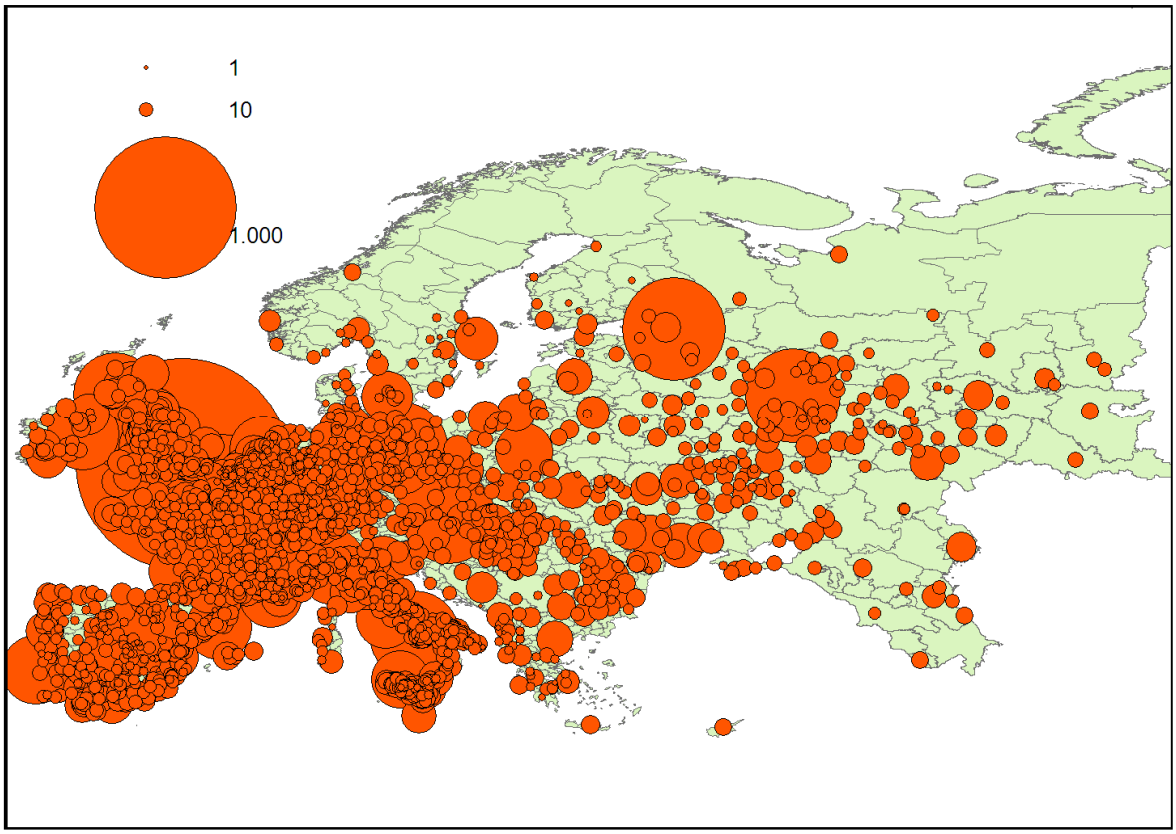
Note : ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses.

Table 3 Market access and literacy, ca. 1930

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable				ln(Literacy)				
ln(MA)	0.56***		0.57***	0.51***		0.77***		0.57***
	(0.000)		(0.000)	(0.000)		(0.000)		(0.000)
ln(Dist. to Lux.)		-0.24***			-0.24***		-0.36***	
		(0.000)			(0.000)		(0.000)	
Constant	-5.72***	0.25***	-5.84***	-5.21***	0.32***	-7.62***	0.72***	-5.78***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Estimation	OLS	OLS	IV	OLS	OLS	Tobit	Tobit	IV
Obs. excluded	Estimated	Estimated	Estimated	None	None	None	None	None
Observations	201	187	186	327	306	327	306	301
Prob. (F-Statistic)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.44	0.38	0.49	0.50	0.44			0.56

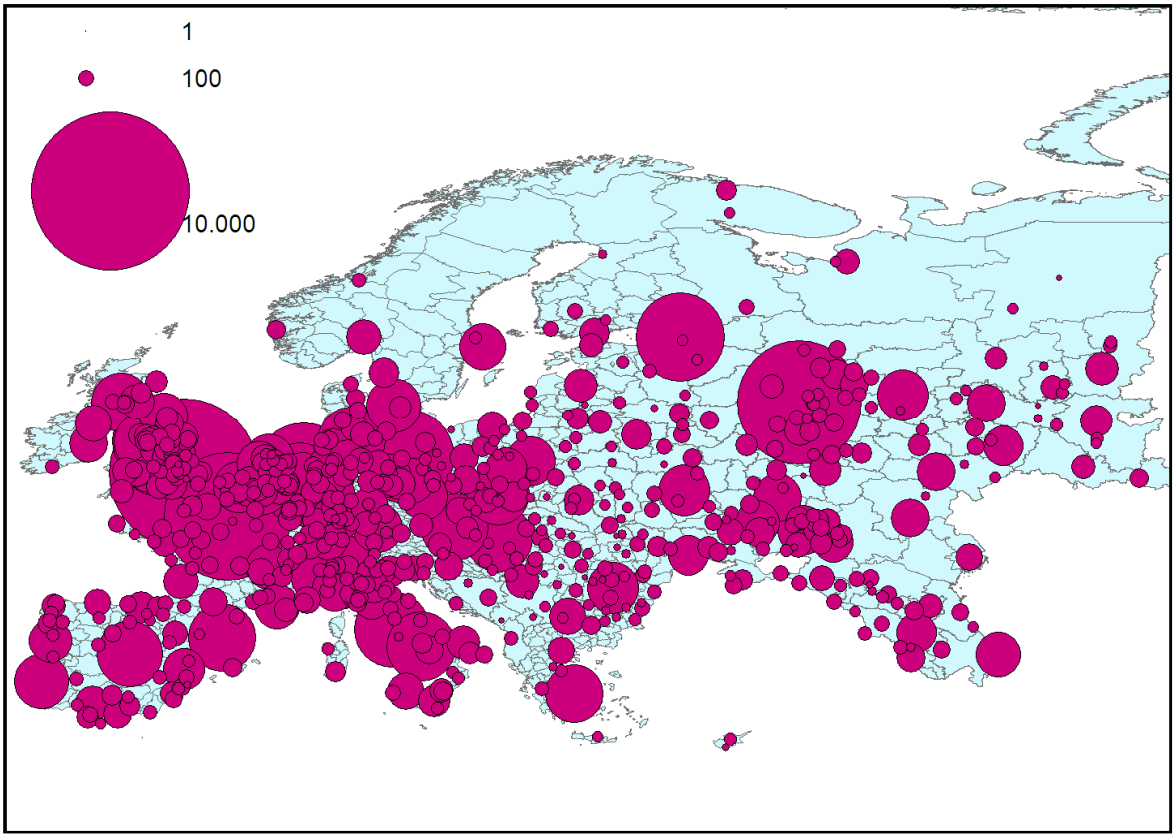
Note : ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses.

Figure 1 Location and size of European cities, 1850



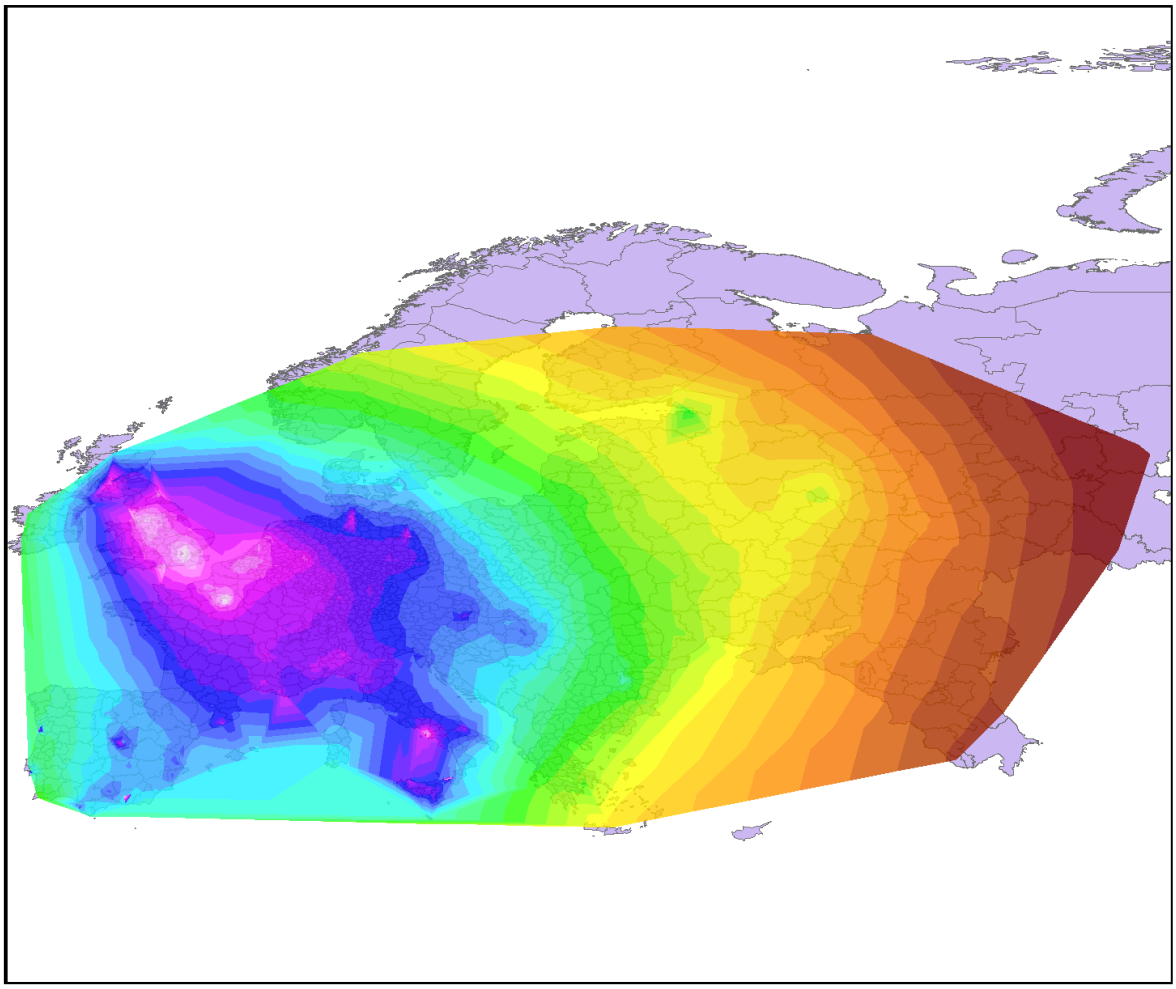
Source: Own graphical presentation of data provided by Bairoch *et al.* (1988). Size of cities is shown in thousand inhabitants.

Figure 2 Location and size of European agglomerations, 1950



Source: Own graphical presentation of data provided by Moriconi-Ebrard (1994). Size of agglomerations is shown in thousand inhabitants.

Figure 3 Population potential in Europe in 1850



Note: Graphical representation using natural breaks (Jenks) with 32 classes. Values decrease from the highest to the lowest value in the following broad order of colours: white, pink, blue, green, yellow, orange and red.
Source: Own calculations, city data provided by Bairoch *et al.* (1988).

Figure 4 ABCC and market access, 1850

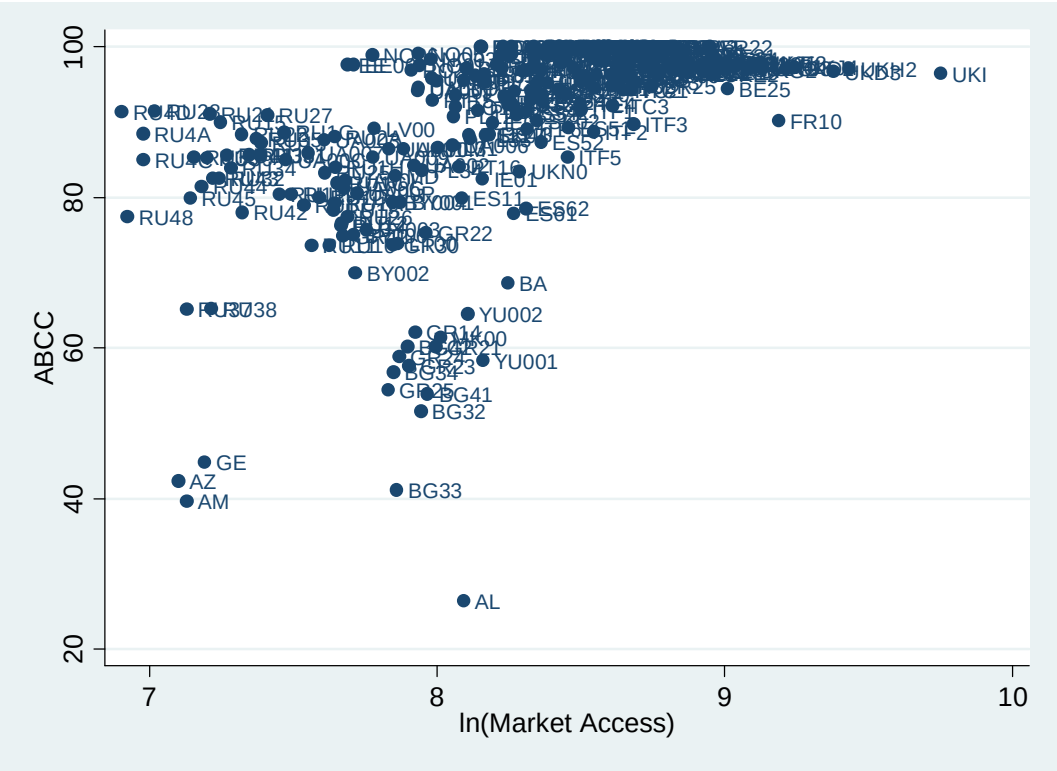
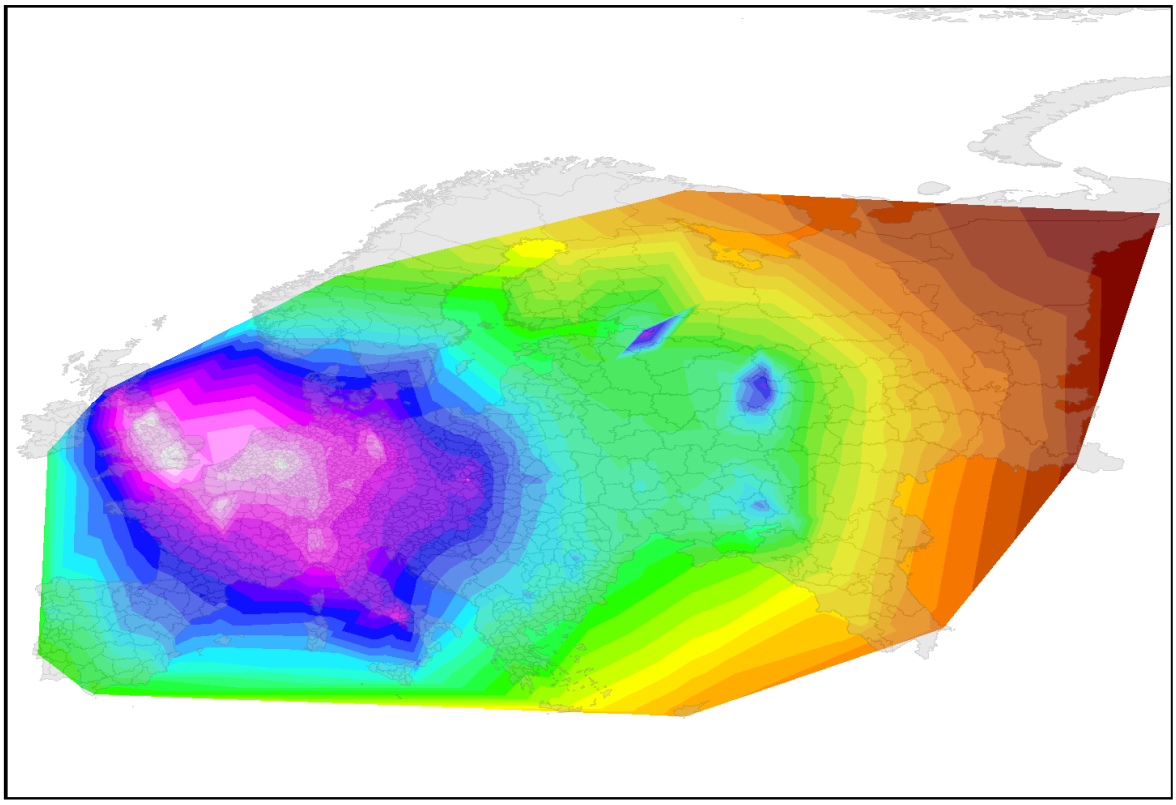
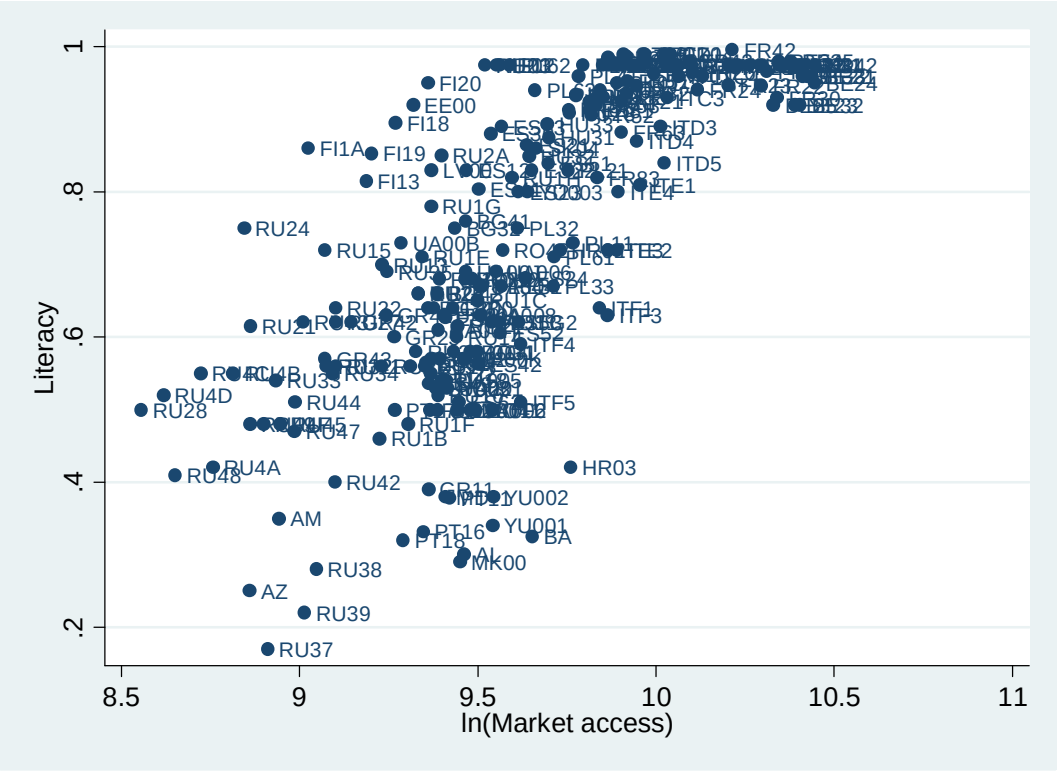


Figure 5 Population potential in Europe in 1950



Note: Graphical representation using natural breaks (Jenks) with 32 classes. Values decrease from the highest to the lowest value in the following broad order of colours: white, pink, blue, green, yellow, orange and red.
Source: Own calculations, data on agglomerations provided by Moriconi-Ebrard (1994).

Figure 6 Literacy and market access, ca. 1930



Appendix

Data

Regional data in 1850 include the following countries (in current borders):

Albania, Armenia, Austria, Azerbaijan, Bosnia-Herzegovina, Belgium, Bulgaria, Belarus, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, France, Georgia, Greece, Croatia, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Moldova, FYROM, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovenia, Slovakia, Ukraine, United Kingdom, Serbia, Montenegro.

Regional data in 1930, without estimated observations, include the following countries (in current borders):

Albania, Armenia, Austria, Azerbaijan, Bosnia-Herzegovina, Belgium, Bulgaria, Belarus, Czech Republic, Estonia, Spain, France, Greece, Croatia, Hungary, Italy, Lithuania, Latvia, Moldova, FYROM, Poland, Portugal, Romania, Russia, Slovakia, Ukraine, Serbia, Montenegro.

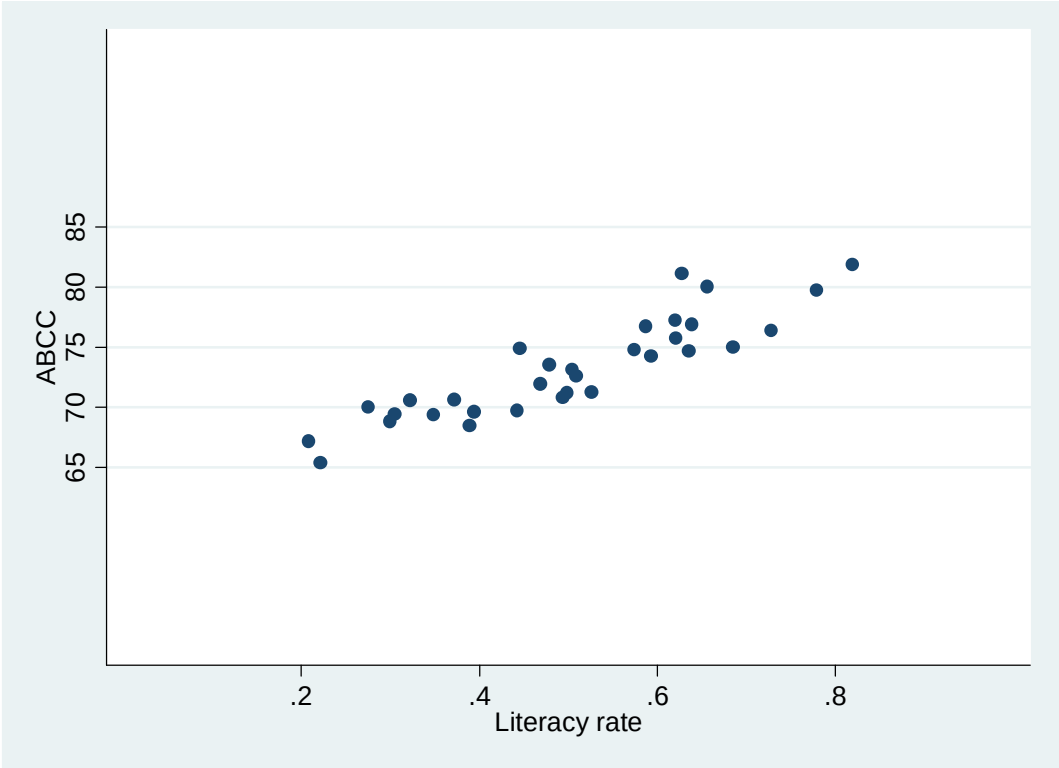
Regional data in 1930, with estimated observations, include the following countries (in current borders):

Albania, Armenia, Austria, Azerbaijan, Bosnia-Herzegovina, Belgium, Bulgaria, Belarus, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, France, Greece, Croatia, Hungary, Ireland, Italy, Lithuania, Latvia, Moldova, FYROM, Netherlands, Norway, Poland, Portugal, Romania, Russia, Sweden, Slovakia, Ukraine, United Kingdom, Serbia, Montenegro.

Numeracy as an indicator for human capital

Numeracy and other human capital indicators are generally closely related. Recent research has provided evidence for this in a range of contexts, including literacy (e.g. A'Hearn *et al.* 2009, Hippe 2012b) and primary school enrolment (e.g., Crayen and Baten 2010). Detailed evidence most closely related to the current study is provided by Hippe (2012b). Hippe analyses numeracy and literacy in the European regions in the 19th century, taking a similar sample used in Hippe and Baten's (2012) and thus our database. For example, considering the case of Irish regions in 1841, the close relationship is quite clear:

Literacy and numeracy in Ireland, 1841



Source: Hippe (2012b).

However, one would not always expect such an almost perfect pattern to emerge. Numeracy measures other human capital characteristics than literacy. Its method of calculation is different from literacy, as literacy is most often provided in historical censuses. In contrast, numeracy is derived from age distributions provided in such censuses but it is not directly stated in the censuses. In this way, it may in some cases have advantages to literacy, as the distribution of literacy is related to the distribution of linguistic (and thus ethnic) groups. The provision of linguistic skills in the predominant state language was often one important aspect of mass education in 19th century Europe. Census questions sometimes only included a lower number of languages to choose from as actually existed in a country. For example, in the Russian Empire it was not deemed possible (and wanted) to include all languages in the census questionnaire, as the number of different languages amounted to the hundreds. Likewise, imagine the case for countries such as France, where a high share of French citizens did not speak French in the middle of the 19th century. The case is even more evident in countries such as Austria-Hungary. While literacy is the standard indicator for the 19th and parts of the 20th century, it may be prone to some biases that numeracy is not. In this respect, the use of two alternative indicators allows us in this study to avoid any conclusions which are solely based on one of these proxies. This provides a higher level of certainty and robustness.

Note that the characteristic age heaping pattern is not a European cultural trait. It can be found similarly in other countries and world regions. For example, Hippe (2012b) also considers microcensus data from developing countries in Africa, Asia and Latin America in the second part of the 20th century. Comparing the derived numeracy values with literacy data, he finds similar tendencies as in historical Ireland, illustrating the close relationship between numeracy and literacy throughout time and independent of cultural characteristics in these considered continents:

Note: bo = Bolivia, br = Brazil, cl = Chile, co = Colombia, ec = Ecuador, in = India, ke = Kenya, mx = Mexico, pa = Panama, tz = Tanzania
Source: Hippe (2012b).