Région et Développement
n° 45-2017
www. regionetdeveloppement.org

Regional human capital inequality in Europe in the long run, 1850-2010

Claude DIEBOLT[*]
Ralph HIPPE[**]

Abstract - Human capital is an important factor for economic and social development, as has been underlined by recent theoretical models. A range of contributions has focused on the international evolution of human capital over the last decades and beyond. However, the regional dimension of human capital in Europe remains insufficiently explored, particularly in a long-run perspective. For this reason, this paper addresses this gap in the literature and highlights the regional evolution of human capital in Europe between 1850 and 2010 by using numeracy, literacy and educational attainment proxies. The results show that intranational inequalities in human capital have always been important and are in a number of cases more important than international differences. However, we find a convergence in literacy and educational attainment rates over the respective time periods.

JEL Classification
N33, N93, O18

Key Words
Human Capital
Regional Development
Inequality
Europe

[*] BETA/CNRS, Université de Strasbourg. cdiebolt@unistra.fr ; claude.diebolt@hu-berlin.de.
[**] European Commission, Joint Research Centre (JRC), Directorate for Growth and Innovation, Human Capital and Employment Unit ; ralph.hippe@ec.europa.eu.
Disclaimer: The views expressed are purely those of the writers and may not in any circumstances be regarded as stating an official position of the European Commission.

INTRODUCTION

Human capital has obtained considerable attention from both researchers and public policy makers recently and in the more distant past. Human capital is often assumed to positively affect a variety of socio-economic factors such as economic development (Lucas 1988, Romer 1990, Galor 2005a, b, Galor 2012), democracy and human rights (McMahon 1999, Sen 1999, Beach 2009).

Nevertheless, many authors concentrate either on the recent development and significance of human capital or on its evolution in history. For example, Gennaioli et al. emphasise in a recent seminal contribution the "paramount importance of human capital in accounting for regional differences in development" in the world today (Gennaioli et al. 2013, p. 105).[1] If human capital has such a striking explanatory power for regional inequalities, then a better understanding of the historical origins of current regional human capital differences appears to be fundamental to comprehend economic development. Are regional inequalities in human capital, then, a product of the modern times or are they a heritage from a distant past? This is a very important issue, as it gives a better understanding of how large regional inequalities can be avoided, which may ultimately threaten the integrity of a country. In addition, it provides an idea about the degree to which educational policy decisions can have an impact of the human capital endowment of a region, and places educational policies into a larger context. The fact that the European Union has witnessed a rise in regional inequalities in recent years makes this issue still more urgent. New insights in this area are also relevant for developing countries which can take lessons from the historical European experience to foster their regional development.

There are already some studies that establish a link between historical and more recent human capital data. However, these studies typically take as the basic unit of analysis either the country level to make international comparisons (e.g., Barro and Lee 2001, Morrisson and Murtin 2009) or they take the regional level to focus on one country (e.g., Felice 2012). In fact, studies that focus on the regional level in Europe and take a long-term approach, bringing together current data with distant historical ones, are rare. When it comes to regional human capital the lack of information is even more striking.

To fill this gap in the literature, this paper explores for the first time the long-term evolution of human capital in Europe at the regional level. More specifically, we show the regional patterns in various education indicators at different points in time. These patterns give a better impression on how regional inequalities in education have changed over time. In addition, we investigate whether regional education measures have converged over time or whether there was a pattern of increasing divergence.

To this end, we construct a new and large database on human capital between 1850 and 2010 from a multitude of sources. More specifically, we employ different indicators for subsequent periods of time. In particular, we use three proxies: numeracy, literacy and educational attainment. Numeracy is measured by the ABCC index (see Hippe and Baten 2012a) and measures very basic calculating skills. Literacy is defined as the ability to read and write, and to have a similar measure for basic levels of educational attainment, we consider the share of individuals who do not have a low level of education (0-2), i.e. who can broadly be considered not

[1] The same is true when considering the city level. For example, Simon and Nardinelli (2002) find that high levels of initial human capital are the drivers for faster city growth in the US between 1900 and 1990.

to be 'early school leavers'[2]. These proxies have some important characteristics in common which make them particularly appropriate for the study of their respective time period. In this way, we can underline the regional evolution of human capital by taking into account the years 1850, 1900, 1930, 1960, 2000 and 2010. In particular, we consider the evolving inequality in human capital, employing several standard measures, such as the coefficient of variation (CV), the Gini coefficient and the Theil index. The use of different indicators at the different points in time does not allow a direct comparison of inequality over time, so that we focus on the common patterns at each point. For a better comparison, we also adapt historical regional boundaries to current NUTS 2 regions. In total, we have between 160 and 340 NUTS 2 regions in our database for the different years.

The results show that intranational inequalities in human capital have been important at the different points in time. Regional differences are in many cases quite persistent and are in a number of cases higher than international ones. Convergence takes place in literacy between 1900 and 1960 and in educational attainment between 2000 and 2010. The inequality measures highlight important variations in inequality between countries throughout time. These findings underline the limitations of cross-country analyses and the need for further human capital research at the regional level in Europe.

The paper is structured as follows. First, we highlight some of the most important contributions in the human capital literature that make long-term comparisons or trace the long-term evolution of human capital in Europe. The second part explains the basic underlying methodology and portrays the different data sources that have been used in this study. Subsequently, the results on the regional evolution of human capital and on the regional inequalities are highlighted. Finally, a conclusion sums up the paper.

1. HUMAN CAPITAL FORMATION IN EUROPE IN THE LONG RUN

Human capital has been emphasised to be a crucial factor to improve the lives of individuals (e.g., Vincent 2000). For example, contrary to the first exogenous growth models, endogenous growth models have stressed the important role of human capital (Lucas 1988, Romer 1990). Human capital enters in some of these models as a separate factor in the production function. Human capital is understood as one form of capital (alongside physical capital and other forms) which allows to use this concept for growth accounting.[3] Furthermore, a long-run view on economic growth has been proposed by Unified Growth Theory (e.g., Galor 2005a, b). Unified Growth Theory highlights that human capital is essential for the creation of long-run growth. However, human capital is a theoretical concept that cannot be measured easily empirically. This is especially true in the long run. Nevertheless, the literature has put forward different proxies for the long-run formation of human capital in Europe. In particular, it is possible to trace its evolution in Europe by the use of numeracy, literacy and book production.

A'Hearn et al. (2009) provide information on numeracy (i.e., the ABCC index) on a number of European countries between 1350 and 1850. They show that there is a general tendency of increasing numeracy values over time. Let us first focus on western and northern Europe. Around 1450, the Netherlands were already more advanced in numeracy than (the more developed) northern Italy. The split between the north and the south of Italy is apparent in their data because the

[2] However, note that the definition is not exactly the same as the one currently used by the European Commission (in particular concerning the included age group). For further information, see section 3.

[3] For some principles of human capital theory see Becker (1962) and Schultz (1962).

southern part of Italy had very low numeracy levels both around 1450 and 1500. Data for southern Italy is lacking for the centuries afterwards but we know from Hippe and Baten (2012a) and Felice (2012) that important differences were still visible at the beginning of the 19th century which became consecutively less pronounced until around the middle of the 20th century (see also Gagliardi and Percoco 2011). According to Felice (2012), there was a renewed (but small) tendency of divergence in the decades after 1960.

The differences in the other western and northern European countries between 1600 and 1850 are less striking. The UK was a numeracy leader in 1700 but other countries such as Denmark, the Netherlands, Belgium, France and Norway reached soon higher numeracy levels. Belgium appears to have been quite rapidly advancing as its catch-up phase was relatively short between 1700 and 1800.

On the other hand, A'Hearn et al. (2009) also show the evolution of many central and eastern European countries. These countries had generally lower numeracy values than their western and northern European counterparts. In Germany, there was a similar divide as in Italy in 1700 but this time it is not a clear geographical criterion but a religious one: the Protestant regions were more advanced in numeracy than the Catholic ones, giving further evidence to theories underlying the positive influence of Protestantism on human capital such as Weber (1958) and Becker and Woessmann (2009). Central European German speaking countries (Germany, Austria, Switzerland) had higher numeracy than countries to their east in 1700. Nevertheless, the latter progressed throughout the time period whereas Protestant Germany fell back between 1750 and 1800, Switzerland becoming the numeracy leader in front of Austria and Poland. This description of numeracy has until now been limited to the country level (with some exceptions) but different projects are underway that will also highlight more regional differences within these countries in numeracy in the near future (see also Juif and Baten 2012). These new projects will considerably improve our understanding on the formation of human capital in the European regions in the very long run.

Second, research on literacy has allowed further insights into European human capital formation. For example, Houston (2001) portrays the evolution of regional male literacy in western Europe from before 1700 until 1970. He defines a threshold (i.e., at least 50 % of males between 20 and 50 years have to be literate) and divides the European regions and countries into different categories. These categories show when a region has surpassed this threshold level. Similar to the evolution in numeracy, Houston indicates that Germanic countries (Germany, Switzerland, the Netherlands, Sweden) were the leaders in literacy, having surpassed the threshold already before 1700. This might also be due to the fact that those countries were at least in part Protestant countries. The south-east of England and the larger Edinburgh areas in Scotland were similarly quick as the aforementioned countries. Moreover, literacy spread gradually to neighbouring regions in Belgium, France and northern Italy as well as to the other regions in Great Britain (except Wales) and Iceland until 1790. Geographical proximity appears to have been a decisive factor in the diffusion of knowledge in general and of literacy in particular.

The pattern is still visible but less striking for the regions surpassing the threshold until 1850. French men from almost all regions became by majority literate during this time period. Exceptions are the Celtic region of Bretagne, some central regions and Corsica. It is possible that the language barrier which separated some of these regions from their French speaking neighbours played a role here. The same suggestion can be made for the late progress in Wales and Ireland. The strongholds of Gaelic speaking regions in Ireland's western part only surpassed the threshold until 1900. However, one has to keep in mind that literacy is

defined by reading and writing a particular language (often the official and not the regional language) and this might have biased the results here. In Spain, the northern-central regions were the literacy leaders. Most other Spanish regions surpassed the limit only until 1900 (except in the south where it took even more time). In the same class fall the north-western regions in Portugal, the remaining regions in Ireland and France as well as the north to central Italian regions. Finally, the last European regions became by a majority (male) literate, except for some southern and northern Portuguese regions which were the last ones to achieve the threshold until 1970.

More generally, there are different databases available for the second half of the 19th century until today. These are in part international databases, including European databases mostly at the national level or sometimes referring to large regions constituting those countries. Some of the most well-known are provided by Banks (1971), Flora (1983), Benavot and Riddle (1988), Barro and Lee (2001), Mitchell (2003), De La Fuente and Doménech (2006), Cohen and Soto (2007) and Morrisson and Murtin (2009) (see an overview in Table 1). One has to add that these databases are not always independent from another. In particular, the most recent ones are in part constructed from earlier databases and take different measurement and correction methods. As can be easily seen, the most popular proxies for the last decades have been educational attainment and years of schooling. The discussions and the intention to improve these databases underline once more the need for more and better data (e.g., Krueger and Lindahl 2001, De La Fuente and Doménech 2006).

Table 1. Databases on international evolution of human capital in the longer term

Authors	*Time period*	*Examples of human capital proxies*
Banks (1971)	1860-1966	Primary, secondary and tertiary school enrolment; literacy; number of books
Barro and Lee (2001)	1960-2000	Educational attainment, years of schooling
Benavot and Riddle (1988)	1870-1940	Primary enrolment rates
Cohen and Soto (2007)	1960-2010	Educational attainment, years of schooling, enrolment rates
De la Fuente and Doménech (2006)	1960-1995	Educational attainment, years of schooling
Flora (1983)	1810-1977	Enrolment rates, number of pupils, number of teachers
Lindert (2004a, b)	1830-2000	Enrolment rates, years of schooling, teachers
Mitchell (2003)	1830-1919	Primary, secondary and tertiary education (number of pupils and teachers)
Morrisson and Murtin (2009)	1870-2010	Years of schooling

Nevertheless, even given these ongoing improvements at the international country level, the regional level is still not adequately represented in Europe. Only for the last 10 to 15 years are comparable regional data available, especially from the European Statistical Office Eurostat. If one wants to go back further in time, data collection becomes much more difficult. This is true in particular if one is interested not only in western but also in eastern Europe. This is surprising given the striking relevance of human capital for the economy and the importance of regional differences in human capital. Accordingly, Cipolla believes that the use of

national averages "in a number of cases [...] conceal internal differences that are as interesting and significant as international variations" (Cipolla 1969, p. 15-16). Therefore, how has human capital in the European regions evolved in the European regions? Have regional inequalities really been important? And if so, have they been as important as international ones (or even more important)? This paper makes a first step to fill this gap in the literature.

2. METHODOLOGY AND DATA

2.1. Measure of human capital

For the comparison of regional human capital in the long run it is important that the employed variables follow some basic common principles. Clearly, this is a difficult task as, theoretically, there are many different possibilities to measure human capital. However, in practice their number is substantially reduced due to lacking data availability in the different European countries. For example, the use of income-based approaches (e.g., using the skill premium) are not possible at the European regional level. Therefore, we use an education-based approach to human capital. More specifically, we use three education variables that may each be considered the representative variable of their respective time period.

The first variable is numeracy, calculated by the use of the age heaping method. It is still a rather recent method but has become a very dynamic research field as evidenced by the number of publications over the last years (e.g., A'Hearn et al. 2009, Manzel and Baten 2009, Crayen and Baten 2010a, Hippe and Baten 2012a, Hippe and Baten 2017). Numeracy is a particular appropriate proxy if the aim is to measure human capital in early periods of human development. This proxy can be employed until the 20th century in many European countries and in some of them even later on. Because regional data for other human capital variables is much more restricted and less broad in geographical coverage, it may be expected that this method will still have more success and more contributions in the future. The details of this method have already been discussed in other publications (see e.g., Hippe and Baten 2012a, Hippe 2012b,). Thus, we do not go into more detail in this paper. Let us only mention that there are characteristic heaping patterns on ages in historical and even in some currently less developed countries' censuses. In particular, a part of individuals did not report their exact age but rounded it on 0 and 5. The most important reason for this was that they did not know and were not able to calculate their age. It can be shown that one can measure numerical capacities by taking advantage of this rounding pattern. In practice, the ABCC Index is defined as :

$$ABCC_{jt} = 125 - 125 \times \left(\sum_{i=5}^{14} n_{5ijt} \Big/ \sum_{i=23}^{72} n_{ijt} \right) \quad (1)$$

where *n* is the number of all observations and *i* is the number of years in region *j* at point in time *t*. Numeracy is used to measure human capital around 1850.

Second, the next variable is literacy. Literacy is measured by the ability to 'read and write'. This proxy has been used for a long time and is still used in many international publications today (see e.g., UNESCO 2005). In short, it measures the reading and writing abilities of individuals as stated to census takers or as filled out in census forms. We take one common definition for literacy that was used in the earlier decades of the 20th century and that is similarly used until today:[4]

[4] For more details see appendix.

$$Literacy_{jt} = \sum_{i=10}^{N} rw_{ijt} \Big/ \sum_{i=10}^{N} n_{ijt} \quad (2)$$

where *rw* is the number of individuals who are able to 'read and write' and *N* is the total number of years of age. Unfortunately, some countries do not collect information on the literacy of the total population, as is the case e.g. for the Scandinavian countries at the turn and the first decades of the 20th century. Hence, they cannot be included in this study. Finally, literacy as the measure of the education of the population is progressively replaced by the level of educational attainment during the 20th century in most countries.

Therefore, the third and last variable is educational attainment. It is one of the standard measures used in today's official publications and has been widely used in the literature on human capital today and in the recent past (and in part beyond) (e.g., Redding and Schott 2003, López-Rodríguez et al. 2005, Breinlich 2006, Rodríguez-Pose and Tselios 2011). It measures the share of individuals that have surpassed a certain educational threshold level, in particular primary, secondary or tertiary education. Clearly, education systems vary importantly throughout Europe and have been subject to changes throughout history. This makes it more difficult to compare educational attainment. Still, it is possible to obtain a level of sufficient standardisation which allows to compute human capital values. This is common practice in international publications in general, comparing different countries in the world, and in publications on Europe and the European Union in particular. Eurostat provides standardised measures for all of its members. A further advantage of taking regions as the unit of analysis as compared to countries is that the regions are bound to the same educational system and have to adhere to the same ruling principles. Therefore, within country comparisons are generally not biased by differences in the education systems. We measure educational attainment in the following way:

$$Notlowedu_{jt} = 1 - \left[\sum_{i=15}^{N} le_{ijt} \Big/ \sum_{i=15}^{N} n_{ijt} \right] \quad (3)$$

where *le* is the number of individuals who have achieved pre-primary, primary and lower secondary education as highest level of education and *n* is the number of all individuals.[5] We have opted for this definition because both numeracy and literacy indicators are proxies of rather basic human capital. In contrast, taking e.g. the share of individuals with tertiary education would clearly be an indicator of more advanced human capital. This measure would be less revealing on the abilities of the overall population. For this reason, it appears more appropriate to choose a proxy for the attainment of rather low education. This proxy captures the basic attainment of the entire population.

[5] This definition is based on the availability of data by Eurostat, referring to ISCED levels 0 to 2. Note that the Eurostat data are derived from the EU Labour Force Survey and refer to the "economically active population", including both employed and unemployed individuals and following the principles set up by the ILO (see e.g., ILO 1982). In contrast, the census data for Russia in 2012 refer to the overall population. Alternatively, it is also possible to take the age range of 25 to 64 years old for the countries provided by Eurostat. Because both age ranges are correlated to 99.5 %, the results do not importantly change when using the alternative age range. We prefer the definition of ages above 15 years because it allows us to include the data on Russia in 2010 and enables easier comparison with literacy.

2.2. Indicators of inequality

Apart from providing some descriptive evidence on regional inequalities, we also use more analytical tools. In particular, we use weighted indexes of dispersion to analyse regional inequalities. More specifically, we use the coefficient of variation (CV) as our regional inequality measure. The CV is defined as:

$$CV = \frac{\sigma}{\mu} \times 100 \tag{4}$$

where σ is the standard deviation of regional human capital values and μ represents the average population-weighted human capital value. This measure is especially appropriate for our study because it is a number without dimensions and thus enables an easy comparison between the countries of our dataset which are characterised by very different mean values. As an alternative standard measure of inequality, we construct population-weighted Gini coefficients (see Jenkins, 1999, 2010 for details of calculation). This standard measure allows us to countercheck our CV results.

For long-term comparisons, although it is clear that each of our human capital variables does not measure exactly the same attributes of human capital, it is important that the variables have some common characteristics to improve their comparability. There are several features that are common or at least similar to all three variables which underline the meaningful use of these variables in the present study.

First, all variables are in some sense representative of the contemporaneous period at which they were collected. More specifically, the problem that arises is that often only one educational variable is provided in a census. In early censuses, no direct educational variable (such as literacy) is provided in many cases, which is why the use of the age heaping method to calculate numeracy values is appropriate. Later on, literacy emerges as a standard educational variable. When literacy levels are relatively high (and thus numeracy levels even more so), censuses often begin to report educational attainment, with illiteracy as the category for the lowest educational attainment (i.e., no educational attainment at all). Therefore, educational attainment has become today a standard variable for education. However, the simultaneous use of literacy as a part of the educational attainment variable does not allow a clear comparison between both 'independent' variables. Such a pattern can be found for a number of countries such as Italy and Russia. In consequence, it appears reasonable to employ the most basic indicator of human capital (i.e., numeracy) at the beginning of the period, then a more advanced one (literacy) and finally a current standard educational variable (educational attainment). Clearly, the use of one variable throughout the period would improve comparability. However, neither numeracy nor literacy can be used after the beginning or the middle of the 20th century and no other variable is available at the regional level for the whole of Europe throughout time. Our use of several proxies is, therefore, currently the best available way to show regional inequalities over a longer time period.

Second, the human capital variables are well correlated with each other. For example, the literature has shown that there is a close correspondence between numeracy and literacy levels in censuses from the US (A'Hearn et al. 2009), developing countries (Hippe 2012a) and Europe (Hippe and Baten 2012, Hippe 2012a). In these studies, it was possible to derive numeracy and literacy values from the same census and for the same set of regions. Moreover, Crayen and Baten (2010) have shown the correlation between historical numeracy and schooling.

Third, all educational variables have mostly been calculated (or are in part directly taken) from official census publications or official surveys. These official documents are generally intended to provide public policy makers and the general public with indications on the state of the population in different contexts. They may be better suited than other rather unofficial documents. For this reason, the common underlying document types underline the methodology common to all variables.

Fourth, some of the variables may be argued to be more or less direct output measures. This is clear for the case of literacy where the ability to 'read and write' is measured. This is also true for numeracy although the output has to be computed from an age distribution. In view of educational attainment, this is also the case for tertiary education. Even though one may consider secondary education not to be a direct output measure, it is still a relevant variable characterised by distinctive regional inequalities. In particular, early school leavers play a significant role here. They are those individuals "who leave education and training with only lower secondary education or less, and who are no longer in education and training" (Council of the European Union 2011, C191/01). The Council of the European Union confirms that this issue is important. It stresses that the reduction in the number of these early school leavers is essential to achieve some of the key objectives of the Europe 2020 strategy (Council of the European Union 2011). Still in 2009, the early school leavers constituted a share of 14.4 %.[6] This arguably still (too) high share illustrates that the successful accomplishment of primary education and parts of secondary education cannot be taken for granted for the whole population until today. In consequence, this has even more so been the case in the past.

Fifth, all numeracy, literacy and educational attainment measures are considering an important part of the total population. In other words, they are not restricted to some particular social groups in society such as military recruits or married couples which are commonly used to approximate literacy by using signature rates. Furthermore, they normally consider both sexes (not only males) in contrast to military recruitment data. Thus, all variables are a measure of the basic human capital of the overall population and are not subject to biases that may arise when only a part of the population is considered.

Sixth, and directly connected to this, one should consider a similar part of the population with regard to age. All variables are commonly defined by the use of a certain age threshold, i.e., they only take into account the individuals above a certain age threshold. Numeracy takes account of the great majority of the individuals that constitute the population of the other two variables. In consequence, all proxies measure the share of individuals that have similar years of ages.

Finally, the seventh and maybe the most important common characteristic for the actual measurement of human capital is the definition of the variable at stake. All three variables are defined by an identical value range which goes from 0 (or 0 %) to 100 (or 100 %). The reason for this is that every variable considers the share of individuals that have some form of education. The rest of the individuals does not have this attribute. In other words, every variable is derived from a binary indicator. In consequence, all variables are subject to the same advantages and disadvantages inherent to share measures (for consequences see also Hippe 2012b). This common measurement framework makes it particularly appropriate to consider numeracy, literacy and educational attainment for estimating regional human capital in Europe from 1850 until today.

[6] Therefore, the goal of the Europe 2020 strategy is to reduce this share to 10 % until 2020 (Council of the European Union 2011).

Given this number of common features, human capital data have been collected from a variety of sources. First, the data on numeracy have been taken from the large database by Hippe and Baten (2012a). Second, literacy data have been added for 1900 and 1960 from the census publications of the different European countries under study (see appendix for more details). Moreover, the data referring to 1930 have been taken from Kirk (1946). Finally, educational attainment data for 2000 and 2010 have been collected from Eurostat (Eurostat 2011). They have been supplemented by census data from Russia in 2010.[7]

Indeed, there are a number of limitations and problems which are related to the use of different datasets. On the one hand, the datasets have to be merged. However, the regional coverage of each database is different. In consequence, only a lower number of regions can be studied when several datasets are considered together. On the other hand, there may be differences in the definition of the variables. For example, the literacy definitions are not always homogeneous across countries and throughout time. Although the differences in the historical data are mostly not large (see appendix), the interpretation of the results should always be considered within this context. The more recent data mostly stem from Eurostat, so that maximum comparability of the data has already been ensured.

Finally, the consideration of a long time scale necessitates the use of a common framework for the classification of regions. Otherwise, the results would lack comparability throughout time. For this reason, we have used the NUTS classification. The NUTS classification is the official European classification which has been developed by the European Union. Clearly, the historical regions of Europe do not always correspond to the current ones. Spain and France are good examples where this is largely the case. Eastern Europe is a quite different matter as empires have broken up and important border shifts occurred during the last 150 years. To this end, the historical regions are adapted as best as possible to fit the current regions. More specifically, we standardise our approach by using NUTS2 regions. NUTS2 regions are for example the *régions* in France or the *Regierungsbezirke* in Germany. In this way, we construct a database with 160 to 340 NUTS 2 regions depending on the considered year. Note that the NUTS 2 level can be a fairly large aggregation level, in particular for smaller countries (for example, Slovenia has only 2 NUTS 2 regions). Thus, the low number of regions could also lead to a high sensitivity in the inequality measures. Given these limitations, the results always need to be taken with some caution.

3. RESULTS

3.1. Evolution of human capital in the European regions, 1850-2010

A first intuition on the overall evolution may be derived from descriptive statistics for all human capital variables. They are given in Table 2. Before 2010, we have the highest number of regions in 1850, taking advantage of the large database by Hippe and Baten (2012a). The number of observations is decreasing during the 20th century until the most recent data for 2000 and 2010. Interestingly, although the indicators and the time periods are very different, the descriptive statistics show that not only the number of regions but also the standard deviation, minimum and maximum values are very similar in 1850 and 2010.

Overall, we may summarise that there is sufficient variation for all variables at all points in time to obtain relevant and pertinent results. The definition of literacy between 1900 and 1960 is (almost) identical so that we may make some general

[7] The Russian census data have been classified according to similar classes as specified in the Eurostat data, i.e. ISCED levels 0 to 2.

comments on its evolution. Note, however, that not always the same countries are included (in particular the data for 1960 refer only to peripheral European countries). Nevertheless, as one would have hypothesised, literacy is progressing over the time period 1900 to 1960 and regional inequalities decrease as literacy approaches its maximum level. Moreover, one may indicate that the number of observations and the distribution of our measure of educational attainment in 2000 are relatively similar to those of literacy in 1930. Data on more regions become available in 2010, but the minimum and mean values show that educational attainment is progressing until the present.

Table 2. Descriptive statistics for the unweighted human capital indicators

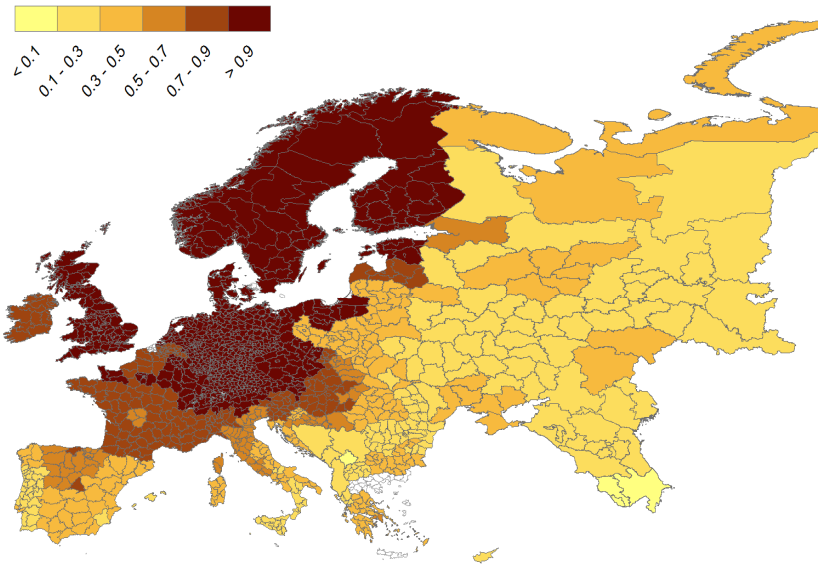
Indicator	year	obs.	mean	sd	min	max
ABCC	1850	304	91.96	11.64	26.38	100.00
Literacy	1900	239	0.57	0.30	0.05	1.00
Literacy	1930	228	0.73	0.21	0.17	1.00
Literacy	1960	169	0.82	0.11	0.59	0.99
Educational attainment	2000	254	0.70	0.16	0.14	0.95
Educational attainment	2010	339	0.77	0.12	0.26	0.97

Let us now turn to the more detailed analysis of the data. Due to the different points in time covered in this paper, we limit our analysis of each point in time to the most important aspects and discuss the most noticeable changes. To begin with the data on the ABCC around 1850, Hippe and Baten (2012a) note that the most important regional differences exist within Bulgaria, Serbia, Spain and Russia (see also the standard deviation in the data appendix). There is also an important north-south difference in Italy and France, to a (much) lesser degree the reverse can be said of Norway. Spain appears to be characterised by a core-periphery pattern. The case in Russia is more complicated: the Caucasus region and the Belarusian regions are the least numerate while in particular Estonia has the highest numeracy values. The latter may be attributed to their historical and cultural ties to the most advanced countries in literacy (and numeracy) in Scandinavia.

In the next step, we consider literacy in 1900. Because literacy did not yet attain its maximum value in many countries, there is more regional variation than in the ABCC. Figure 1 shows regional literacy rates in 1900 (see also the data in the appendix). To our knowledge, it is the first time that such a map has been produced for the turn of the century. Thus, it constitutes in itself a contribution to the existing literature. It shows that the regions with the highest literacy rates were located in central and northern Europe. In addition, the Scandinavian countries, the Netherlands, Germany, Luxembourg and Switzerland can be assumed to have literacy rates above 90 %, so that they would fall into the highest literacy category. These are the countries which have historically been the leaders in literacy (see Houston 2001). It is also a similar argument to Kirk's (1946) reasoning for 1930. Therefore, we argue that this choice is well justified.[8] Furthermore, the map gives an overall impression of a core-periphery model, with central Europe at its core.

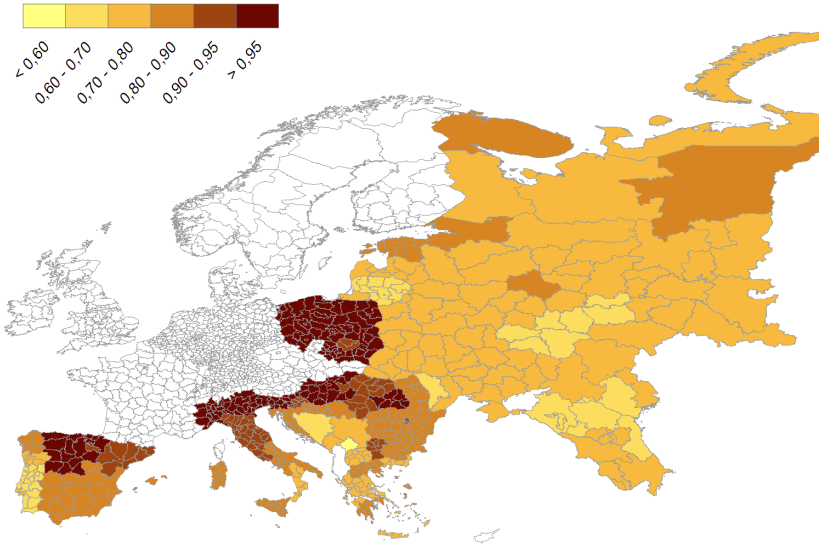
[8] But, of course, further data would provide more information about regional variation even within those advanced countries.

Figure 1. Literacy, c. 1900 (NUTS2)



Note: Data for historical Germany, Denmark, Finland, Luxembourg, Netherlands, Norway and Sweden are not available. For mapping purposes, their literacy rates have been estimated to be above 90 %.

Figure 2. Literacy, c. 1960 (NUTS2)



Note: Only national data available for Belarus and Ukraine, data unavailable for Polish Opolskie region.

Still, there are a number of regions that perform better or worse than one would expect from their pure geographical location. For example, the Kosovo region appears to have had comparatively low literacy levels. On the other hands, regions such as Santander in Northern Spain appear to outperform other regions. Capital regions, such as Madrid in Spain, Attika in Greece and St. Petersburg in Russia, have the tendency to have higher literacy. The importance of national administrations and the relevance of literacy in an urbanised environment are probably important drivers of this pattern. Note also that Cisleithania (i.e., the Austrian part of Austria-Hungary) has a very high range in literacy to that time (from about 0.2 to almost 1). Regional inequalities are also persistent in a number of other countries between 1850 and 1900. This is the case e.g. for Spain, Hungary and Italy. Less but still important variation is apparent for the regions belonging to the Ottoman Empire, Portugal and the Russian Empire.

Advancing to 1930 (not shown), many results of the previous points in time are confirmed, underlining the persistence of human capital inequalities. In 1960, the literacy scale already indicates that regional variation has considerably decreased (Figure 2). For this reason, note that the number of considered countries is limited to less developed regions in western, eastern and south-eastern Europe. The highest regional variation can be observed in Spain, Italy and Serbia. Portugal has the lowest average literacy rates in Western Europe, even lower than most regions in the East. The northern and larger urban regions of St. Petersburg, Moscow and Murmansk are positive outliers in Russia. The patterns are in many ways similar to the ones found in 1930 and 1900. Although the Austrian-Hungarian Empire dissolved after the First World War, its geographical limits are still visible more than forty years later. Similarly, the historical regional differences in Yugoslavia are still perceivable. Path dependency is clearly one important feature that emerges from this analysis. In general, we also see that the numeracy and literacy patterns broadly reflect the available previous information in the literature (such as Houston 2001) but provide a larger and in part more detailed information, in particular also highlighting disparities in Eastern Europe.

Taking educational attainment in 2000 (not shown), we can now include much more countries. Still, Portugal is at the lower bound of educational attainment. This highlights the historical continuity of Portuguese low human capital performance in a European comparison. Its capital, Lisbon, has significant higher levels of educational attainment. Nevertheless, Lisbon is still at the very bottom of the educational attainment of other countries (except Malta which has still lower educational attainment). The highest average values come from the Czech Republic and Slovakia which constituted one country until 1992. Other former Communist countries have also high educational attainment levels, such as Poland and Hungary. Given our indicator, within-country regional inequalities are relatively small when compared to literacy data in 1900 or 1930. For example, Italy's regional educational disparities are relatively small. Still the highest inequalities are found in Spain. Finally, the current situation in 2010 (not shown) is similar to that in 2000 although some countries have been added. Russian disparities are relatively small but historical literacy leaders are once more significant positive outliers (Moscow, St. Petersburg, Murmansk). In general, one can state that educational attainment has been rising during the first decade of the 21st century. In particular, the heritage of educational policies in former Communist countries in central Europe is clear given the high values for educational attainment. Nevertheless, regional inequalities are striking until the present, underlining the importance of the regional level for academic research and policy makers.

Figure 3. Growth in regional human capital

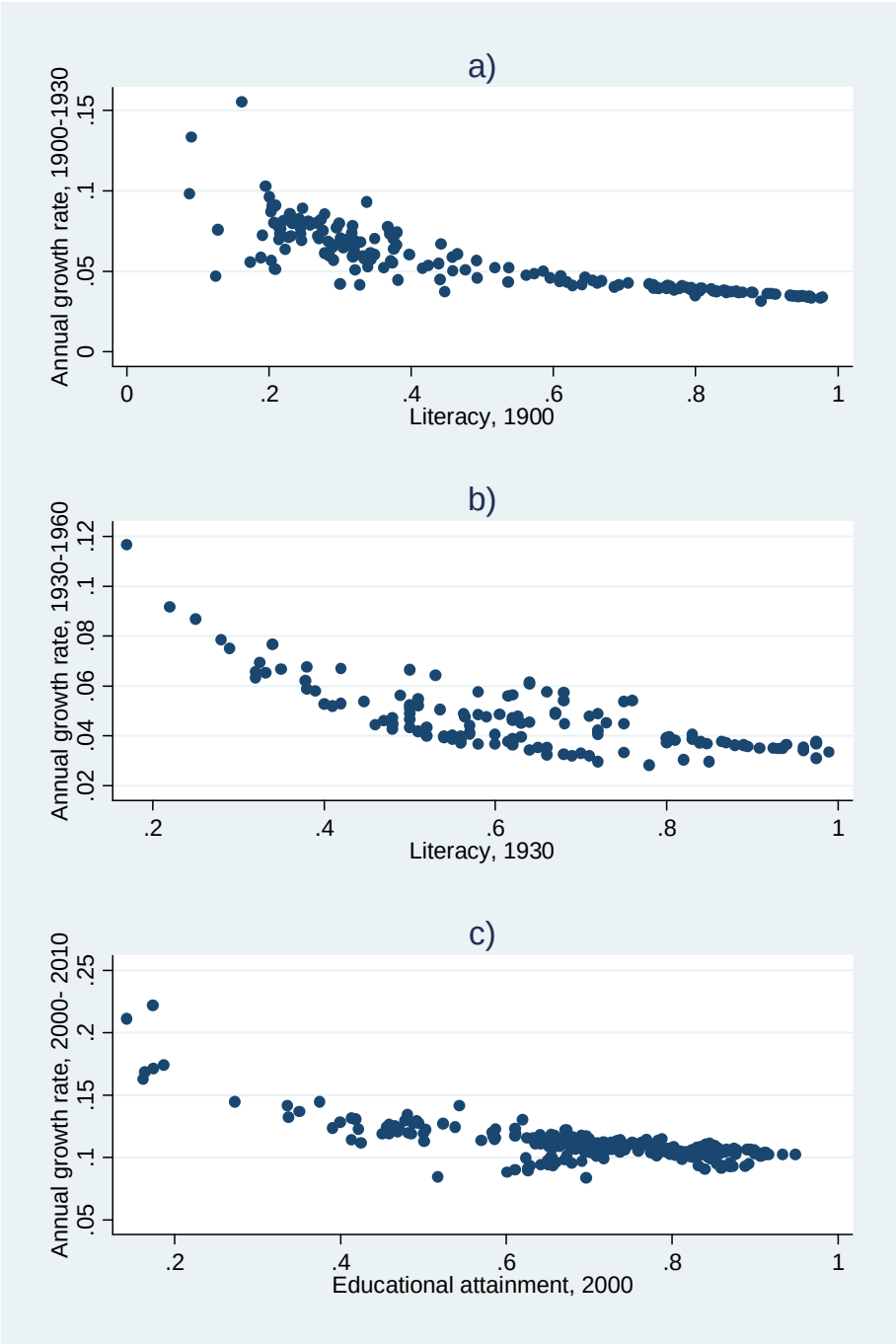
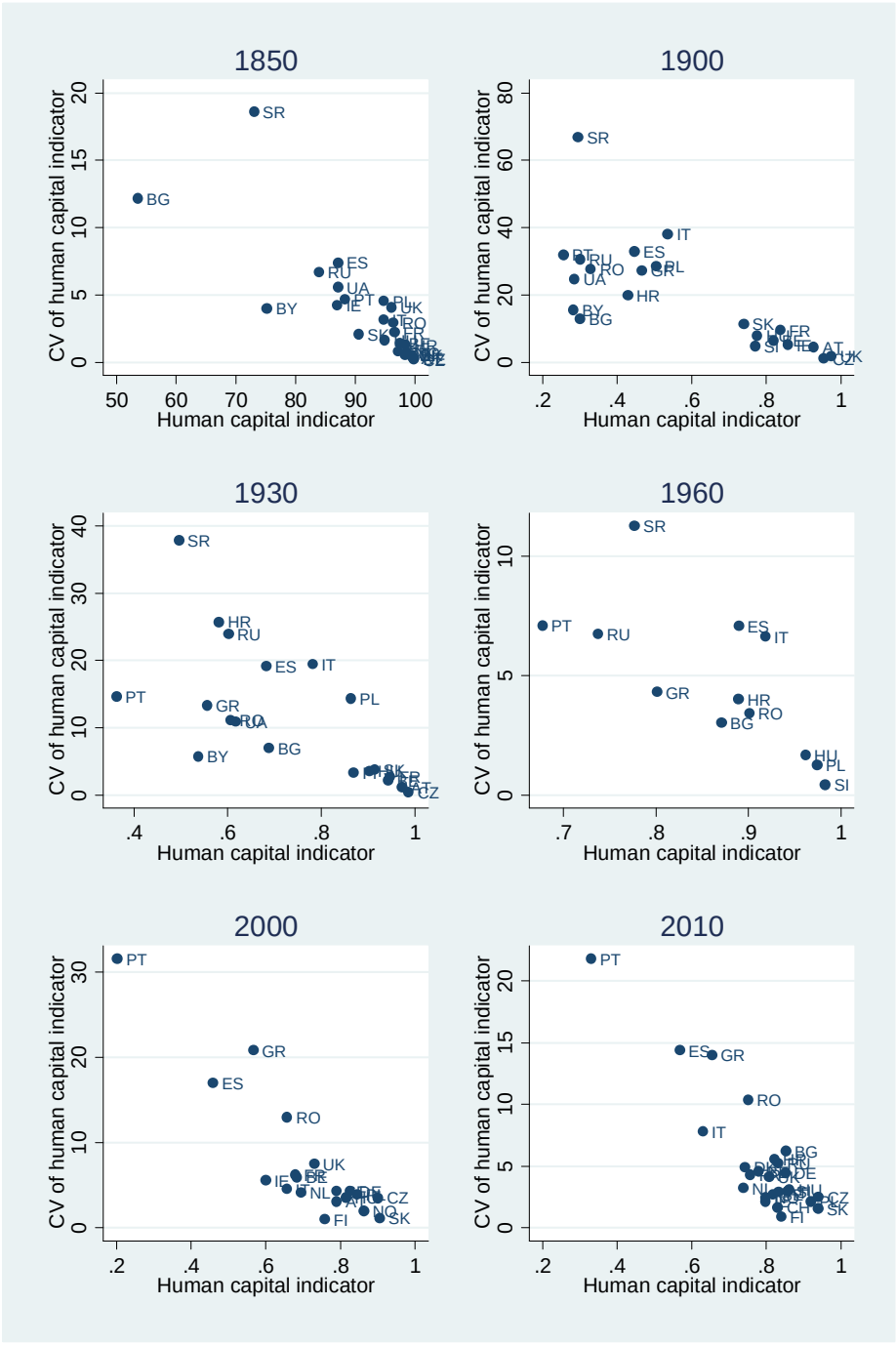


Figure 4. Regional inequality and level of human capital



After these first comparisons in the long run, we should take a more specific look at the changes from one point of time to another. Given the fact that our three human capital variables are not directly comparable, we emphasise the changes in literacy between 1900 and 1960 and in educational attainment between 2000 and 2010.

In Figure 3 we consider the initial literacy rates in 1900, 1930 and 2000 and compare them with the subsequent growth over the next years. We define the annual growth rate as:

$$\Delta HC_{j,t-T} = \frac{HC_{j,t}/HC_{j,T}}{t-T} \tag{5}$$

where HC is the human capital variable, ΔHC is the annual growth rate in human capital, j a region, t is the latter point in time and T is the initial point in time. Note that the dataset is importantly reduced in 1960, so that we have fewer observations when considering 1960.

The part on the top of the figure (a) shows this case for literacy in 1900 and growth in literacy between 1900 and 1930. We find a clear convergence scheme where those regions with the highest literacy rates (and thus closest to the maximum literacy limit) grow the least, and those with low literacy rates grow much stronger. Still, there are a number of regions that underperform in this respect, generating noise at the lower end of the distribution. These regions come particularly from the Russian Empire, Yugoslavia and Portugal. Among others, these low performances may be derived from the effects of the Russian Revolution, the First World War, the construction and formation of new regions and countries. But also positive outliers can be found. For example, the highest growth is found for Albania and the second highest for Armenia in the top-left corner, being at the same time at the very end of the literacy scale together with Azerbaijan. The most important negative outlier in the bottom-left corner is the Russian region Dagestan, an ethnically mixed region located north of Azerbaijan with an important share of Muslims. One could hypothesise that the efforts of the Russian central authorities to increase literacy may have not been very effective because it had only limited power in this region.

The same plot for 30 years later (but a lower number of observations) does not appear to show so important outliers from the general pattern (b). Clearly, many regions have now higher literacy rates, so that one might also expect to find important regional differences in growth rates particularly at middle values of the literacy scale. In particular, we find that regions whose growth rates were higher than others between literacy levels of 40 and 80 % came from Bulgaria and Romania.

Finally, we find again a similar pattern for current educational attainment between 2000 and 2010 (c). The lowest attainment values with the highest growth refer to Portuguese regions.

We can also use econometric tools to check whether we find convergence. Similar to Zhang and Li (2002), we regress the natural log human capital in the initial year on the average annual human capital growth rate (see Table 3). In columns 1 and 2 we include all regions. In all cases, the coefficient is negative and significant at the 1 % level. Convergence is evident. Interestingly, the negative coefficient is smaller in column 1 than in column 2, indicating that convergence is faster in the period 1930 to 1960 than in the period 1900 to 1930. However, we have to take into account that the number of observations is importantly reduced in 1960. Therefore, we check our previous results in columns 3 and 4 by only including those regions that are available at all concerned points in time (i.e., 1900, 1930 and

1960). However, the basic finding remains. We can also control for country effects, including dummies for today's countries. In that case, the negative coefficient for 1900 to 1930 becomes smaller and the one for 1930 to 1960 larger, indicating that the finding of increasing speed of convergence in the latter period remains robust (not shown). Thus, the findings suggest that β-convergence (i.e., convergence in average literacy levels across regions) is taking place, and convergence is faster in the second period than in the first one.

Table 3. Convergence in regional human capital

Variables	(1) 1900-1930	(2) 1930-1960	(3) 1900-1930	(4) 1930-1960	(5) 2000-2010
ln(HC 1900)	-0.0296***		-0.0281***		
	(0.000)		(0.000)		
ln(HC 1930)		-0.0323***		-0.0322***	
		(0.000)		(0.000)	
ln(HC 2000)					-0.0429***
					(0.000)
Constant	0.0319***	0.0298***	0.0326***	0.0299***	0.0943***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Regions	All	All	Reduced	Reduced	All
Observations	178	141	118	118	254
R-squared	0.73	0.66	0.60	0.62	0.70

*Dependant variable: ΔHC. p-values in parentheses; *** p<0.01, ** p<0.05, * p<0.1.*

In column 5, we indicate the coefficient for the period 2000 to 2010. However, as the human capital variable is different, we cannot compare this result for those obtained for literacy. Still, convergence is taking place more rapidly for this variable during the more recent years.

3.2. Intranational inequality

The evolution of inequality within a country is the next crucial aspect to take into account. The complete results when using the CV are shown in Table 4. As an alternative standard measure of inequality, it is also possible to construct population-weighted Gini coefficients.[9] The results are quite similar to those obtained for the CV (not shown). Therefore, the use of the Gini coefficient validates the results obtained by using the CV. In consequence, for simplicity we will simply refer in the following to the CV. It appears important to emphasise that our methodology, that is the use of different human capital proxies, does not allow a direct comparison of CVs and Ginis over time. Such a comparison is not intended here nor wanted. Instead, our focus is on the order of countries with regard to regional inequalities.

It is clear that Serbia was marked by very high regional differences. These differences may result from the fact that Serbia was created from regions that formerly belonged to countries with very different levels of development, e.g., Austria-Hungary and the Ottoman Empire, and that we consider Kosovo to be still part of Serbia. Although regional inequalities decrease until 1960, Serbia was not able to relinquish these differences.

Furthermore, Portugal appears to have been characterised by important regional differences throughout history even though this country is rather small and homogeneous in cultural terms. This finding underlines once more that regional

[9] See Jenkins (1999/2010) for details of calculation.

variation can be striking even for small countries. Other countries with high CVs are in particular Italy, Spain, Greece and Russia. Still, these CVs have different causes in each case. For example, Italy shows striking north-south differences over the centuries. More specifically, the north has relatively high levels of human capital while the south lags behind. The core-periphery pattern in Spain has led to similar important regional disparities. The Greater Athens region in Greece is not only the capital regions of the country but in this way also the administrative, economic and educational centre, giving it an advance to other Greek regions. Finally, although we are only considering the European part of Russia, this is still a huge area, comprising different ethnical and cultural groups, so that regions do not develop their human capital at an identical pace throughout time.

Table 4. CVs at different points in time

Country	1850	1900	1930	1960	2000	2010
AT	0.323	4.529	1.190		3.089	2.920
BE	1.411	6.481	2.241		5.971	4.571
BG	12.175	12.961	7.031	3.037		6.223
BY	3.999	15.565	5.734			
CH	0.781					1.663
CZ	0.218	1.119	0.406		3.528	2.516
DE	0.295				4.306	4.454
DK	0.542					4.929
ES	7.412	32.949	19.180	7.078	17.021	14.409
FI			3.310		1.017	0.909
FR	2.230	9.471	2.703		6.252	4.314
GR		27.285	13.333	4.320	20.836	13.993
HR	1.299	19.853	25.744	4.020		5.571
HU	1.635	7.992	3.499	1.684	3.586	3.097
IE	4.234	5.209			5.594	2.137
IT	3.153	38.138	19.448	6.654	4.542	7.806
NL	0.544				4.135	3.226
NO	0.718				1.971	2.450
PL	4.571	28.613	14.337	1.264	3.888	2.128
PT	4.659	31.880	14.667	7.109	31.568	21.818
RO	2.932	27.712	11.125	3.423	12.952	10.363
RU	6.689	30.643	23.920	6.752		5.227
SE					4.321	2.685
SI	0.518	4.801		0.437		2.926
SK	2.095	11.324	3.782		1.145	1.585
SR	18.641	66.775	37.854	11.277		
UA	5.592	24.735	10.880			
UK	4.096	1.804			7.498	4.142

We also further investigate the relationship between the regional inequality within one country and the national average of human capital. Presumably, approaching the maximum level of human capital will lead to a decrease in regional inequality. In Figure 4, we show that this idea is largely the case. The years 1850, 1960, 2000 and 2010 appears to show this tendency relatively clear. However, the year 1900 shows a different pattern. In fact, there is the tendency to increasing inequality until a level of about 50 % which then decreases with higher levels of

literacy. The only country that does not follow this pattern is Serbia. Serbia is again a clear outlier, for the reasons noted above. Accordingly, apart from Serbia, Italy has the highest CV in this sample, being also close to 50 %. This pattern is only observable in this case because it is the only year that we have a relevant number of national averages of a human capital indicator of below 50 %. All other years are concentrated on levels higher than 50 %. Indeed, this potential inverted U-shaped form of regional inequality in human capital appears to indicate a Kuznets curve of regional human capital inequality. A 'Kuznets curve of human capital inequality' has been recently suggested by Morrisson and Murtin (2013). Their methodology is quite different from ours as they do not consider regional inequality but inequality in the distribution of education (as measured by different levels of schooling), taking the national level and calculating the returns to education. Therefore, this tentative evidence could suggest that we find a Kuznets curve also for regional inequality when using our methodology.

Table 5. Negative differences between European and national Ginis and CVs

Country	Year	Gini difference	CV difference
BG	1850	-0.016	-1.189
ES	2010	-0.007	-0.243
GR	2000	-0.017	-2.229
GR	2010	-0.004	
HR	1930	-0.001	-2.354
PT	2000	-0.038	-12.960
PT	2010	-0.031	-7.652
RU	1930	-0.005	-0.530
SR	1850	-0.035	-7.655
SR	1900	-0.051	-18.620
SR	1930	-0.042	-14.464

Finally, we want to address the question whether intranational inequality may sometimes be higher than international inequality. To this end, we calculate the average CV and Gini at the European level and compare this to the national CVs and Ginis. The results are shown in Table 5. Apart from Greece in 2010, where a negative difference exists only for Ginis, the countries that have higher intranational inequality than international inequality are the same when using Ginis and CVs. These countries are Bulgaria, Spain, Greece, Croatia, Portugal, Russia and Serbia. Here, we only consider the whole sample of European countries. Comparing regional inequalities to smaller groups of countries (e.g., a number of neighbouring countries), we would certainly find even more countries where regional inequalities are higher than international ones. Still, these examples emphasise that regional inequalities can be quite significant and in some cases they may even be larger than between-country differences.

These results suggest that the study of regional data can offer important insights that get lost in pure cross-country analyses. In a longer run perspective, regional inequalities appear to be a natural phenomenon which, however, can be tackled (at least to some extent) if appropriate policy decisions are taken.

CONCLUSION

This paper has traced the long-run evolution of human capital in the European regions between 1850 and 2010. Human capital is an important factor that has to be considered in a variety of setups because it affects economic and social developments. The role of human capital has particularly been stressed by parts of the economic growth literature in the last decades, from Endogenous Growth Theories (Lucas 1988, Romer 1990) until the recent contributions by Unified Growth Theory (e.g., Galor 2005a).

We have constructed a new and large database from a variety of sources. The data from different points in time have allowed to construct a dataset that covers at least an important share of all European countries in 1850, 1900, 1930, 1960, 2000 and 2010. To our knowledge this is the first time that it has been possible to show and analyse the regional evolution of human capital during such a long time period for such an important number of countries located in the European continent. To this end, three different proxies have been used: numeracy, literacy and educational attainment. More specifically, numeracy is measured by the ABCC Index, literacy by the share of individuals able to 'read and write' and educational attainment by the share of employed individuals having attained an education level above lower secondary education. We show that this choice is not arbitrary but that the inherent characteristics of these variables make them appropriate for the purposes of this study.

For a general overview of the evolution of human capital in the long run, we have presented some of the literature that has contributed evidence on human capital in a longer historical perspective. More precisely, we have put forward the evolution of numeracy, literacy and book production.

After this point of departure, we have traced the evolution of human capital in the European regions between 1850 and 2010. The most striking result is that regional differences have been significant in many European countries in the past and the present. The persistence of regional human capital differences is striking in many cases. Literacy (1900 to 1960) and educational attainment (2000 to 2010) rates have shown to be converging over time. In addition, we measure regional inequality by the coefficient of variation and the Gini coefficient. Both show that some countries have been much more unequal than others. The high regional differences within countries, which sometimes can be higher than intercountry differences, underline that cross-country analyses miss an important part of the human capital story. For this reason, more research needs to be done at the regional level in Europe. This paper has done a further step into this direction. The new evidence is important both for academics and policy makers because it contributes to the understanding of long-term developments which are important for the present and the future.

In particular, it further emphasises the need to better understand the territorial dimension of human capital. At this moment, regional information on human capital is mostly only available in form of different levels of educational attainment. However, in this study we were able to use historical output measures in the form of numeracy and literacy, and it would be more than appropriate to have similar measures for today. In other words, large scale surveys such as PISA, TIMSS, etc. could be used to a greater degree for regional studies. So far, regional data are mostly limited to a few countries and most studies in this area only consider one or two countries. Further data collection at the regional level would increase the knowledge upon which policy makers can effectively take the appropriate policy decisions.

In addition, the existence of long-term educational structures indicates that policy reforms have to be designed in a way that they take into account a range of 'deep' institutional, cultural and economic structures that may be related to educational outcomes. This means that structural changes may be needed which may not always be an easy task – but the rewards can be large, both in terms of educational improvements and, potentially, also long-run growth (see also Hanushek and Woessmann 2016).

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.

Banks, A. S. (1971). Cross-Polity Time-Series Data, Cambridge: MIT Press.

Barro, R. J. and J.-W. Lee (2001). International Data on Educational Attainment: Updates and Implications, *Oxford Economic Papers*, 53 (3): 541-563.

Beach, M. J. (2009). A Critique of Human Capital Formation in the U.S. and the Economic Returns to Sub-Baccalaureate Credentials, *Educational Studies: A Journal of the American Educational Studies*, 45(1): 24-38.

Becker G. S. (1962). Investment in human capital: a theoretical analysis, *Journal of Political Economy*, 9-49.

Becker, S. O. and L. Woessmann (2009). Was Weber wrong? A Human Capital History of Protestant Economic History, *Quarterly Journal of Economics* 124 (2): 531-596.

Benavot, A. and P. Riddle (1988). The Expansion of Primary Education, 1870-1940: Trends and Issues, *Sociology of Education* 61 (3): 191-210.

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

Cohen, D. and M. Soto (2007). Growth and Human Capital: Good Data, Good Results, *Journal of Economic Growth*, 12: 51-76.

Council of the European Union (2011). Council Recommendation of 28 June 2011 on policies to reduce early school leaving, 2011/C 191/01, last accessed on 15 August 2014. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2011:191:0001:0006:EN:PDF

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.

De La Fuente, A. and R. Domenech (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?, *Journal of the European Economics Association*, 4: 1-36.

Eurostat (2011). Economically active population by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000), data online.

Felice, E. (2012). Regional convergence in Italy, 1891-2011: testing human and social capital, *Cliometrica*, 6 (3): 267-306.

Flora, P. (1983). *State, Economy, and Society in Western Europe 1815-1975, Vol. I*, Frankfurt: Campus Verlag.

Gagliardi, L. and M. Percoco (2011). Regional disparities in Italy over the long run: the role of human capital and trade policy, *Région et Développement*, 33: 81-105.

Galor, O. (2005a). From Stagnation to Growth: Unified Growth Theory, In: Aghion, P. and S. N. Durlauf (eds.), *Handbook of Economic Growth*, Volume 1A, 171-293, Amsterdam: North Holland.

Galor, O. (2005b). The demographic transition and the emergence of sustained economic growth, *Journal of the European Economic Association*, 3 (2-3): 494-504.

Galor, O. (2012). The demographic transition: causes and consequences, *Cliometrica*, 6: 1-28.

Galor, O. Moav, O. and D. Vollrath (2009). Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence, *Review of Economic Studies*, 76: 143-179.

Gennaioli, N., La Porta, R., Lopez-de-Silanes F. and A. Shleifer (2013). Human capital and regional development, *Quarterly Journal of Economics*, 105-164.

Hanushek, E. A., and L. Woessmann (2016). Knowledge capital, growth, and the East Asian miracle. *Science*, 351(6271): 344-345.

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

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

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

Hippe, R. and J. Baten (2017). Geography, Land Inequality and Regional Numeracy in Europe in Historical Perspective, Working Paper.

Houston, R. (2001). Literacy, in: Stearns, E. (ed.). Encyclopedia of European Social History, Vol. 5, Detroit: Charles Scribner's Sons, 391-406.

ILO (1982). Resolution concerning statistics of the economically active population, employment, unemployment and underemployment, adopted by the Thirteenth International Conference of Labour Statisticians, online, last accessed 15 August 2014. http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087481.pdf

Jenkins, S.P. (1999/2010). INEQDECO: Stata module to calculate inequality indices with decomposition by subgroup, Statistical Software Components S366002, Boston College Department of Economics, revised 19 Apr 2001, online, last accessed 15 August 2014, available at: http://ideas.repec.org/c/boc/bocode/s366007.html

Juif, D. and J. Baten (2012). On the Human Capital of Inca Indios before and after the Spanish Conquest. Was there a "Pre-Colonial Legacy"?, *Explorations in Economic History*, 50 (1): 227-241.

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

Krueger, A. B. and M. Lindahl (2001). Education for Growth: Why and for Whom?, *Journal of Economic Literature*, 39: 1101-1136.

Lindert, P. (2004). *Growing Public*, Cambridge: Cambridge University Press.

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

Lucas, R. (1988). On the Mechanics of Economic Development, *Journal of Monetary Economics* 22 (1): 3-42.

Manzel, K. and J. Baten (2009). Gender Equality and Inequality in Numeracy: The Case of Latin America and the Caribbean, 1880-1949, *Revista de Historia Económica* 27 (1): 37-73.

McMahon, W. W. (1999). *Education and Development: Measuring the Social Benefits*, New York: Oxford University Press.

Mitchell, B. R. (2003). *International Historical Statistics: Europe 1750-1993*, New-York: M. Stockton Press.

Morrisson, C. and F. Murtin (2009). The Century of Education, *Journal of Human Capital*, 3 (1): 1-42.

Morrisson, C. and F. Murtin (2013). The Kuznets curve of human capital inequality: 1870–2010, *The Journal of Economic Inequality*, 11(3): 283-301.

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 .

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

Romer, P. (1990). Endogenous Technological Change, *Journal of Political Economy* 99 (5): 71-102.

Schultz, T. W. (1962). Reflections on investment in man, *Journal of Political Economy*, 1-8.

Sen, A. (1999). *Development as Freedom*, New York: Anchor Books.

Simon, C. J. and C. Nardinelli (2002). Human capital and the rise of American cities, 1900–1990, *Regional Science and Urban Economics*, 32 (1): 59-96.

Subrahmanian, R. (2005). Gender equality in education: Definitions and measurements, *International Journal of Educational Development*, 25: 395-407.

UNESCO (2005). *EFA Global Monitoring Report 2006*, Paris: Graphoprint.

Vincent, D. (2000). *The Rise of Mass Literacy*, Cambridge: Polity Press.

Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism*, New York: Charles Scribner's Sons.

Woessmann, L. (2003). Specifying Human Capital, *Journal of Economic Surveys*, 17 (3): 239-270.

Zhang, J. and T. Li (2002). International Inequality and Convergence in Educational Attainment, 1960–1990, *Review of Development Economics*, 6 (3): 383-392.

APPENDIX

Country abbreviations

abbreviation	country
AL	Albania
AM	Armenia
AT	Austria
AZ	Azerbaijan
BA	Bosnia-Herzegovina
BE	Belgium
BG	Bulgaria
BY	Belarus
CH	Switzerland
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GE	Georgia
GR	Greece
HR	Croatia
HU	Hungary
IE	Republic of Ireland
IS	Iceland
IT	Italy
LT	Lithuania
LU	Luxembourg
LV	Latvia
MD	Moldova
ME	Montenegro
MK	FYROM
NL	Netherlands
NO	Norway
PL	Poland
PT	Portugal
RO	Romania
RU	Russia
SE	Sweden
SI	Slovenia
SK	Slovakia
SR	Serbia
UA	Ukraine
UK	United Kingdom

Data for 1850

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

Data for 1900

Country (Census year)	Source
Albania (1918)	Preliminary dataset "Albanische Volkszählung von 1918", entstanden an der Karl-Franzens-Universität Graz unter Mitarbeit von Helmut Eberhart, Karl Kaser, Siegfried Gruber, Gentiana Kera, Enriketa Papa-Pandelejmoni und finanziert durch Mittel des Österreichischen Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Special thanks to Siegfried Gruber for providing the data.
Austria (Cisleithania) (1900)	K. K. Statistische Central-Commission (1903). Oesterreichische Statistik. Ergebnisse der Volkszählung vom 31. December 1900, 2. Band, 2. Heft, Wien, Kaiserlich-Königliche Hof- und Staatsdruckerei.
Bosnia-Herzegovina (1910)	Mayer, M. (1995). Elementarbildung in Jugoslawien (1918-1941), München, R. Oldenburg Verlag.
Belgium (1900)	Statistique de la Belgique (1903). Population. Recensement général. 31 décembre 1900, Bruxelles, Typographie-Lithographie A. Lesigne.
Bulgaria (1900)	Principauté de la Bulgarie (1906). Résultats généraux du recensement de la population dans la principauté de Bulgarie au 31 décembre 1900, 1-ère livraison, Sophia: Imprimerie "Gabrovo".
Cyprus (1911)	Mavrogordato, A. (1912). Cyprus. Report and general abstracts of the census of 1911, London: Waterlow & Sons Ltd.
France (1906)	Statistique Générale de la France (1908). Résultats statistiques du recensement général de la population effectué le 6 mars 1906, Paris.
Greece (1907)	Royaume de Grèce (1909). Résultats statistiques du recensement général de la population effectué le 27 octobre 1907, Tome I, Athènes: Imprimerie nationale.
Hungary (Trans-leithania) (1900)	Magyar Statisztikai Közlemények (1907). A magyar szent korona országainak 1900. Evi. Népszámlálása. Harmadik rész. A népesség részletes leirása. Budapest: Pesti Könyvnyomda-Részvénytársaság.
Ireland (1901)	Census of Ireland, 1901 (1902). Part II. General Report, Dublin: Brown & Nolan Ltd.
Italy (1901)	Ministero di agricultura, industria e commercio (1907). Anuario statistico italiano 1905-1907, Fascicolo Primo, Roma: G. Bertero e C.
Montenegro (1900)	MacKenzie Wallace, D. (2006). A short History of Russia and the Balkan States, Elibron Classics, Adamant Media Corporation.
Portugal (1900)	Ministerio dos negocios da fazenda (1906). Censo Da Populacao Do Reino De Portugal No 1. De Dezembro De 1900, Vol. II, Lisboa: A Editora.
Romania (1899)	Royaume de Roumanie (1905). Résultats définitifs du dénombrement de la population (décembre 1899), Bucarest, Eminesco.
Russian Empire (1897)	издание центрального статистического комитета министерства внутренних (1899-1905). первая всеобщая. переписъ населения, российской империи, 1897 г., с.-петербург. Various tombs.
Serbia (1906)	Direction de la Statistique d'Etat du Royaume de Serbie (1908). Annuaire statistique du Royaume de Serbie, Onzième Tome, Belgrade, Imprimerie de l'Etat du Royaume de Serbie.
Spain (1900)	Dirección general del Instituto geográfico y estadístico (1903). Censo de la Poblacion de Espana, según el Empadronamiento hecho en la Península é Islas adyacentes en 31 de diciembre de 1900, Tomo II and Tomo III, Madrid: Imprenta de la Dirección general del Instituto geográfico y estadístico.
United Kingdom (1901)	Hechter, M. (1976). U.K. County Data, 1851-1966 [computer file]. Colchester, Essex: UK Data Archive [distributor]. SN: 430, http://dx.doi.org/10.5255/UKDA-SN-430-1. Although all efforts are made to ensure the quality of the materials, neither the original data creators, depositors or copyright holders, the funders of the Data Collections, nor the UK Data Archive bear any responsibility for the accuracy or comprehensiveness of these materials.

Note: Age definitions are as follows: Italy = 6+; Austria, Bosnia-Herzegovina = 7+; Spain = 8+; Albania, Belgium, Bulgaria, France, Greece, Ireland, Portugal, Russian Empire = 10+; Hungary, Romania, Serbia = 11+; Cyprus = 15+; United Kingdom = unavailable (a comparison of the included data for Ireland with the source for Ireland as listed above has revealed similar overall results, so that an age definition of 10+ can reasonably be assumed). Age definitions are either directly given in the publication or have been linearly estimated from available age definitions in order to be as close as possible to the standard definition of ages above 10 years. The various age definitions as calculated here may possibly not significantly affect the final results. For example, the percentage change of using a 5+ instead of a 10+ definition is below 1 % in Ireland.

Data for 1930

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

Data for 1960

Country (Census year)	Source
Bulgaria (1956)	централно статистическо управление при министерския свет (1960). преброяване на населението в народна република българия на 1. XII. 1956 година, общи ресултати, книга II, софия: държавно издателство "наука и изкуство".
Greece (1961)	Royaume de Grèce (1968). Résultats du recensement de la population et des habitations effectué le 19 mars 1961, Vol. III, Athènes: Office nationale de Statistique.
Hungary (1960)	Központi Statisztikai Hivatal (1962). 1960. Évi népszámlálás, Budapest, Allami Nyomda.
Italy (1961)	Istat (2012). Serie Storiche, Tavola 7.1.1, online, last accessed 3 August 2012, http://seriestoriche.istat.it/fileadmin/allegati/Istruzione/tavole/Tavola_7.1.1.xls.
Poland (1960)	Glówny Urzad Statystyczny (1960). Biuletyn statystyczny. Spis powszechny z dnia 6 grudnia 1960 r., Ludnosc. Gospodarstwa domowe, Wyniki ostateczne, Seria "L", various issues, Warszawa.
Portugal (1960)	Instituto nacional de Estatística (1960). X Recenseamento Geral Da Populacao, Tomo III, Lisboa: Sociedade Tipográfica.
Romania (1956)	Republica Populara Romîna (1961). Recensamîntul Populatiei din 21 Februarie 1956, Rezultate Generale, Bucuresti, Direcţia Centrala de Statistica.
Spain (1960)	Instituto nacional de Estadística (1969). Censo de la Poblacion y de la Viviendas de Espana, según la Inscripción realizada el 31 de diciembre de 1960, Tomo III, Madrid: I.N.E. Artes graficas.
USSR (1959)	Demoscope (2012). Всесоюзная перепись населения 1959 года. Таблица 7. Распределение населения по возрасту и уровню образования. РГАЭ. Ф.1562 Оп. 336 Д.1591-1594, online, last accessed 8 August 2016, http://demoscope.ru/weekly/ssp/rus_edu_59.php. Demoscope (2012). Всесоюзная перепись населения 1959 года. Таблица 2,5. Распределение всего населения и состоящих в браке по полу и возрасту. РГАЭ. Ф.1562 Оп. 336 Д.1535-1548. Российская Государственная библиотека, отдел "литературы ограниченного пользования", online, last accessed 8 August 2016, http://demoscope.ru/weekly/ssp/sng_mar_59_r.php.
Yugoslavia (1961)	Statisticni urad Republike Slovenije (2012). Popis prebivalstva 1961, Prebivalstvo, staro 10 let ali več, po spolu, starosti in pismenosti, online, last accessed 8 August 2016, http://www.stat.si/publikacije/popisi/1961/1961_2_40.pdf.

Note: Age definitions are as follows: Italy = 6+; Hungary, Poland, Portugal, Spain = 7+; Bulgaria, Romania = 8+; Greece, USSR, Yugoslavia = 10+. Age definitions are either directly given in the publication or have been linearly estimated from surrounding available age definitions to be as close as possible to the standard definition of ages above 10 years.

Data for 2000 and 2010

All countries except Russia: Eurostat (2011). Economically active population by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000), online, last accessed 22 June 2016, lfst_r_lfp2acedu.
Russia: всероссийская перепись населения о россии языком цифр (2012). Население по уровню образования по субъектам Российской Федерации, online, last accessed 12 October 2016: http://www.perepis-2010.ru/results_of_the_census/tab8.xls.

Les inégalités régionales de capital humain en Europe de 1850 à 2010

Résumé - Les modèles théoriques et empiriques récents montrent que le capital humain est un facteur fondamental pour le développement économique et social. Cependant, la dimension régionale du capital humain en Europe n'a encore été que peu étudiée, en particulier dans une perspective de long terme. Le but de cet article est ainsi focalisé sur l'évolution régionale du capital humain en Europe entre 1850 à 2010. Pour mesurer le capital humain, nous utilisons des indicateurs de capacités numériques, d'alphabétisation et de niveau d'éducation. Les résultats montrent que les inégalités intra-nationales de capital humain sont importantes et, dans quelques cas, davantage marquées que celles observées au niveau international. En revanche, il semble qu'il y ait une convergence au niveau de l'alphabétisation et des niveaux d'éducation.

Mots-clés
Inégalités régionales
Capital humain
Europe
Histoire économique