



**Bureau
d'économie
théorique
et appliquée
(BETA)**
UMR 7522

Documents de travail

« The long-run impact of human capital on innovation and economic development in the regions of Europe »

Auteurs

Claude Diebolt, Ralph Hippe

Document de Travail n° 2016 – 31

Juin 2016

Faculté des sciences économiques et de gestion

Pôle européen de gestion et
d'économie (PEGE)
61 avenue de la Forêt Noire
F-67085 Strasbourg Cedex

Secrétariat du BETA

Géraldine Del Fabbro
Tél. : (33) 03 68 85 20 69
Fax : (33) 03 68 85 20 70
g.delfabbro @unistra.fr
www.beta-umr7522.fr



The long-run impact of human capital on innovation and economic development in the regions of Europe

Claude Diebolt¹ and Ralph Hippe²

Abstract

Human capital is supposed to be an important factor for innovation and economic development. However, the long-run impact of human capital on current innovation and economic development is still a black box, in particular at the regional level. Therefore, this paper makes the link between the past and the present. Using a large new dataset on regional human capital and other factors in the 19th and 20th century, we find that past regional human capital is a key factor explaining current regional disparities in innovation and economic development.

Keywords: Human Capital, Economic Development, Innovation, Regions, Europe.

JEL codes: I25, N90, O18, R11.

¹ BETA/CNRS, Université de Strasbourg; cdiebolt@unistra.fr, claudio.diebolt@hu-berlin.de.

² London School of Economics and Political Science, Grantham Research Institute on Climate Change and the Environment; r.hippe@lse.ac.uk.

1. Introduction

Economic development is one of the predominant research areas in economics. Many theories have been developed to better understand the causes and consequences of economic development and growth. For example, some of the most important fundamental factors for long-run growth are the quality of institutions (e.g., North 1981, Acemoglu *et al.* 2005) and geography and naturally given geographical conditions (e.g., Diamond 1997, Engerman and Sokoloff 2000). Approximate causes of growth include income inequality (e.g., Alesina and Rodrick 1994, Persson and Tabellini 1994), land inequality (e.g., Galor *et al.* 2009) and human capital accumulation (Galor and Moav 2002, Glaeser *et al.* 2004). For instance, an increase in human capital may induce a rise in the number of innovative entrepreneurs and products, thus indirectly spurring economic development through the channel of innovation. In fact, the crucial role of innovation for economic development and growth has been underlined by a large literature in this area (e.g., Solow 1957, Romer 1986, Lucas 1988). Nevertheless, the long-run implications of human capital on innovation and economic development need further research because this issue has only been touched upon in few contexts (e.g., Baten and van Zanden 2008). Therefore, the question remains whether pre-existing human capital is important for the creation of long-run development.

Thus far, most of the studies in this area only take a national perspective by focusing on countries. However, regional differences in human capital may be at least as important as national ones (e.g., Cipolla 1969). The use of regions allows to overcome the inherent problems of cross-country analyses and may explain why some regions are richer than others. In particular, human capital may play a crucial role in regional development. In fact, in their recent seminal paper Gennaioli *et al.* show the “paramount importance of human capital in accounting for regional differences in development” (Gennaioli *et al.* 2013, p. 105).

But is the effect of human capital also persisting? Their analysis is limited to current data and cannot evaluate any longer run influence of human capital on regional outcomes. We aim at assessing this aspect in this paper. Therefore, we analyse the long-run impact of human capital on innovation and economic development at the regional level in Europe. To our knowledge, this is the first paper that takes this long-run regional approach at the European scale, contributing a new spatio-temporal dimension to the existing literature.

Combining a range of databases for the first time, we employ a new and large dataset in our analysis. First, this dataset includes data on human capital levels between 1850 and 2010 for many European regions and countries. Second, the database also comprises relevant current data on innovation and economic development. More specifically, we measure current innovation by patents per million inhabitants and the level of economic development by GDP per capita. Finally, we add historical socio-economic control variables that stem from a number of different sources. These historical control variables include the share of agricultural employment, population density, infant mortality, fertility and marital status. We also include dummy variables for former Communist countries in Eastern Europe and control for capital regions.

Regions are coded according to the European Union's NUTS classification throughout time. In other words, we adapted the historical European regions to the current NUTS system to directly compare the historical to the current data. In total, we have up to 265 NUTS 2 (or corresponding) regions in our database at a point in time. In this way, we are able to analyse the relationship between human capital, innovation and economic development in a regional and long-run perspective.

More specifically, using standard OLS regression models we regress current regional innovation and economic prosperity measures on a range of historical variables at

different points in time. Our baseline specification considers historical explanatory variables in 1930, the year in which we have the maximum number of variables. The results show that historical human capital is a significant determinant of today's regional levels of innovation and economic development in Europe. In particular, literacy has a significant influence on current patent applications per capita and GDP per capita. We employ a number of specifications to check the robustness of our results. Among others, supplementary results for 1850 (using age-heaping based numeracy), 1900 and 1960 (using literacy) confirm our findings. Therefore, our results suggest that historical human capital has important persisting effects on economic development.

The paper is structured as follows. First, we present the relevant literature on the relationship between human capital, innovation and economic development in Europe. Then, we discuss the employed methodology, the underlying data and our econometric strategy. Finally, we show the current relationship between human capital, innovation and economic development and analyse the long-run relationship between historical human capital, current innovation and economic development. The last section concludes.

2. Literature

Human capital may directly affect economic development and growth or indirectly, in particular through the generation of technology. According to Acemoglu and Autor (2012), there are several channels through which human capital may affect technological progress. Firstly, they stress that the individuals with the highest talents may contribute to technological progress by the use of their human capital if they have the necessary access to educational facilities. These individuals have probably the most important impact on technological progress. Secondly, the workforce in more general terms may affect technology, first, due to the externalities derived from human capital and, second, because

human capital alters and increases the incentives to invest more in technological progress. For example, it is possible that a technology is only sufficiently profitable if there are enough workers who have the necessary skills. Finally, technological progress may be influenced by the workforce's mix of skills and human capital.

In general, the importance of human capital was already considered in early works by Smith and Marshall (see Demeulemeester and Diebolt 2011). However, it took much longer for human capital to emerge as a key factor for economic growth. In fact, the most important contributions were developed from the middle of the 20th century onwards. In particular, Becker (e.g., see Becker 1964) is widely acknowledged as a founder of human capital theory, stressing that human capital increases the productivity of workers. Similarly, Arrow (1962) highlights the effect of experience on technical change. In addition, Nelson and Phelps (1966) emphasise that human capital is also important for implementing and adopting new technologies. Later on, Schultz (1975) argues that workers are better able to cope with changes in the economic structure and handle new technologies if they have more human capital.

Around the beginning of the 1990s emerged new theoretical advances. An extension of the original Solow growth model (i.e., the human-capital augmented Solow model) was presented by Mankiw *et al.* (1992). It explicitly includes human capital as a factor in the Cobb-Douglas production function. Another kind of growth models, the endogenous growth models, was initiated by Romer (1986) and Lucas (1988). The former focuses on technological change and the latter on human capital accumulation. The aim is to endogenise the various factors which may lead to economic growth in the model. Overall, these models consider human capital to be an important driver for economic growth. They have also stimulated further research, generating another branch of

Schumpeterian growth models (Aghion and Howitt 1992, 1998, 2006) that model the idea of creative destruction through innovation.

Finally, the newest contribution in the area of human capital theory and economic growth are the Unified Growth models (e.g., Galor and Weil 2000, Galor and Moav 2002, Galor 2005, 2012). Their aim is to explain economic development in the (very) long run. In these models, human capital is attributed a crucial role for the creation of economic growth.

All in all, these different theories show that human capital is an important driver for economic development and growth. Still, there has been some controversy about this issue over the last decades. In fact, Demeulemeester and Diebolt (2011) refer to several alternating waves of optimism and scepticism on the relevance of human capital to generate growth since the Second World War. The contributions by authors such as Solow (1956), Mincer (1958), Schultz (1961) and Becker (1964) led to the consensus in the 1950s and 1960s that education makes an important contribution to economic growth. In contrast, the 1970s were more marked by scepticism in a time of economic downturn. The new important theoretical contributions of the 1990s (Lucas 1988, Romer 1990) reinvigorated once again the case for human capital. These optimistic ideas were supported by different empirical studies (e.g., Barro 1991, Mankiw *et al.* 1992, Barro and Lee 1993) but also more critical voices appeared such as Benhabib and Spiegel (1994) and Pritchett (2001). Measurement error may account for some of these results (Krueger and Lindahl 2001). Thus, Sianesi and Van Reenen conclude in their literature survey in 2003 that “as a whole we feel confident that there are important effects of education on growth” (Sianesi and Van Reenen 2003, p. 197). In addition, the more recent studies by, e.g., De La Fuente and Doménech (2006), Cohen and Soto (2007), Goldin and Katz (2008) and Ciccone and Papaioannou (2009) show the crucial impact of human capital on growth. The key

contribution of cognitive skills (including numeracy and literacy skills) is further highlighted by Hanushek and Woessmann (2008).

The literature on the impact of human capital and innovation on economic development and growth in the European regions is also large (e.g., Fagerberg *et al.* 1997, Rodríguez-Pose and Crescenzi 2008, Sterlacchini 2008, Cuaresma *et al.* 2012). For example, Badinger and Tondl (2003) investigate whether human capital and innovation (as measured by patent applications) have a significant impact on the growth rates of Gross Value-Added per capita in 128 regions between 1993 and 2000. Both the relative patent applications and higher education variables are shown to have a significant impact. However, medium levels of education are not significant which highlights that economic growth in Europe's 'knowledge-driven' economies is boosted by the highest form of educational attainment. Moreover, Sterlacchini (2008) finds that human capital (in the form of higher education) and a region's knowledge base have a significant and positive impact on economic growth in twelve EU15 countries between 1995 and 2002. Cuaresma *et al.* (2012) use a dataset including 255 EU regions to analyse which of their 48 potential determinants are significantly explaining economic growth between 1995 and 2005. Two of their most important results are that capital regions grow faster than other regions and that human capital (i.e., higher education) is a robust determinant of economic growth. Finally, Gennaioli *et al.* (2013) construct a database of 1569 regions from more than 100 countries to disentangle the determinants of regional development. Considering a broad range of geographical, institutional, cultural and human capital variables, they find that human capital is the single most important factor for regional development. Thus, these different studies show that human capital is a crucial determinant of economic growth and economic development in the European regions and in the world today.

But what do we know about its long-term impact in the world in general and in Europe in particular? There have been some studies which shed some light on the question whether historical human capital and technology matter for today's economies. For instance, Comin *et al.* (2010) take a long-run perspective and show that there is a strong relationship between technology in 1500 AD and current GDP per capita as well as technology adoption in the world. Madsen (2008, 2010) shows that the growth effects of human capital are important at the country level in OECD countries over the last hundred or so years, underlining the predictions of Schumpeterian growth models. These findings suggest that historical factors may be important for the explanation of current or recent economic levels.

We advance this line of research by focusing on regions instead of countries in a European perspective. Using regions instead of countries considerably sharpens the picture. Countries may be composed of regions which do not share a common linguistic, ethnical or cultural identity. Regional differences may thus be very high. However, this information is lost in country comparisons. Aggregated country averages may hide the fundamental forces operating at more disaggregated levels. For example, cross-country analyses cannot disentangle national institutional effects on economic outcomes. Therefore, we analyse whether there are persisting long-run effects of human capital on innovation and economic development, using regional historical human capital and current innovation and economic development data.

3. Methodology and data

Human capital, innovation and economic development are rather large and vague ideas whose measurement has to be specified in greater detail. The human capital data used in this study come from different sources. First, we employ the new and large database

created by Hippe (2014) which traces human capital between 1850 and 2010. From this database, we use the years 1850, 1900, 1930 and 1960 to follow the evolution of human capital. Human capital is proxied by numeracy (ABCC) in 1850 and by literacy (ability to read and write) in 1900, 1930 and 1960.

Both numeracy and literacy indicators may be considered appropriate for their respective time period. Before 1900, literacy data are not available for many European countries. Even in 1900 a range of countries do not consider literacy in their censuses. This is the case for e.g. Scandinavian countries such as Denmark or Sweden but also for Germany, Switzerland or the Netherlands. In general, these are countries where basic reading and writing skills can be considered almost universal. They had their own specific reasons to refrain from this question in the census. For example, the Swiss administration considered that a sufficient literacy level was already attained in 1860, as the corresponding 1860 census documents highlight (Statistisches Bureau 1862). According to the census materials, military data had shown that 93 % of recruits were able to read and write in the Bern region and even 100 % of recruits were literate in the Solothurn region already at the middle of the 19th century. Similarly, the Netherlands had already very high literacy levels if one considers recruitment data: only 15 % of recruits were illiterate (not or only unsatisfactorily able to read and write) in 1857/1858 (Statistisches Bureau 1862).

These examples highlight the very high levels in literacy which existed in (probably all of) the countries where literacy was not asked in the census at the end of the 19th century. For this reason, it appears more suitable to use another indicator for the earliest point in time. Numeracy as proxied by the age heaping method is the appropriate choice because, first, it is closely correlated to literacy (Hippe 2012a). Second, numeracy is – as literacy data later on – directly derived from censuses. Third, it refers broadly to the same population (the entire population, excluding certain age groups). This allows a better

comparison of both indicators. Taking military data from recruits would not allow to take the major parts of the population into account but only a very small selected group: men, in military service, of rather younger age and limited to a defined small age range. Moreover, regional data are often not available.

In consequence, numeracy is the appropriate indicator which is also available for almost all European regions around 1850. Numeracy is measured by the age heaping method which has been used in an increasing number of recent publications (A'Hearn *et al.* 2009, Manzel and Baten 2009, Crayen and Baten 2010, Hippe 2012b, Hippe and Baten 2012). The method takes advantage of the fact that in historical censuses there is a heaping phenomenon on ages particularly ending on 0 and 5. One can show that individuals were not able to calculate their own age, so that they did not report their exact age but only a rounded age.

The deviation from the ideal age distribution (where all ages are represented by the same share) can be employed to create an index measuring numeracy. This index has originally been the Whipple index (WI) but has recently been improved by the ABCC Index (see A'Hearn *et al.* 2009). This index has the same value range as literacy (0 to 100 percentage points or simply points) which makes comparisons much easier.

Therefore, we employ the ABCC Index also in this study. It is defined as

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

where i is the number of years, j is a region, t is the point in time (with $t = 1850$) and n is the number of individuals.

Second, literacy was the standard education variable around the turn of the 20th century and the first half of the 20th century in many European countries. Illiteracy had to be eradicated – this was a common tenor in all European countries. Success, however, was

quite different not only in these countries but also *within* these countries (see also Figure 1). For this reason, a completely literate population was not achieved in many European countries in 1900 and still in 1960 illiterates were more or less common in many European countries. This fact underlines our methodology to use literacy as our human capital indicator for the period. After 1960 one may presume that the ability to read and write is more or less attained by the entire population so that other education variables have to be used. We define literacy as

$$Literacy_{jt} = \frac{\sum_{i=10}^N rw_{i,jt}}{\sum_{i=10}^N n_{i,jt}}, \quad (2)$$

where rw is the ability to read and write, N is the total number of years and t is the point in time (with $t = 1900, 1930, 1960$). The age definition is the standard contemporary definition.

Furthermore, innovation is difficult to be measured statistically. One standard way is to take the number of patent applications or grants (e.g., Acs *et al.* 2002, Diebolt and Pellier 2009, Diebolt and Pellier 2012). In addition to patent applications, other variables that are used to measure innovation include investments in R&D (e.g., Cohen and Levinthal 1989), changes in productivity (David 1990, Von Tunzelmann 2000), bibliometrics (Andersen 2001) and data on (international) expositions and fairs (Moser 2005). Patent statistics have certain setbacks; for example, organisational changes or know-how cannot be patented and not all patented products become innovations (Griliches 1990). Nevertheless, patents are generally considered to be the best indicator (e.g., Cantwell 1989, Anderson 2001) and are most frequently employed (Diebolt and Pellier 2009), in particular for the past. Therefore, we use patent applications per million inhabitants to the European Patent Office (EPO) as our indicator of innovation. The regional data come from Eurostat (2014).

Lastly, the level of economic development is measured in a standard way by GDP per capita (in PPS) as presented by Eurostat (2014).

We use scatter plots and regression models to analyse the relationship of regional human capital, innovation and the level of economic development. For the influence of historical human capital on current innovation and economic development, we employ standard OLS regression frameworks which are formulated in the following way:

$$\ln(\text{Patents}/c_j) = \beta_0 + \beta_1 H_j + X_j + \varepsilon_j \quad (4)$$

$$\ln(\text{GDP}/c_j) = \beta_0 + \beta_1 H_j + X_j + \varepsilon_j, \quad (5)$$

where $\ln(\text{Patents}/c)$ is the number of patents per million inhabitants (in logarithmic terms), $\ln(\text{GDP}/c)$ is GDP per capita (in PPS and in logarithmic terms), H is the human capital indicator, X are other explanatory variables, j is a region and ε are the unexplained residuals.

X is composed of different variables which may have an influence on economic development. Our baseline specification considers X (and H) in 1930 because we have the maximum number of variables for this point in time. Thus, in 1930 the explanatory variables are total fertility, marital status, population density, the share of individuals not dependent on agriculture, infant mortality, a dummy for capital regions, a dummy for the newer EU regions and country dummies. There is a large literature showing that fertility can have an important effect on growth (e.g., Becker 1981, Barro and Becker 1989, Becker *et al.* 1990, Galor and Weil 1996, 2000, Galor 2012). According to the quantity-quality trade-off theory, parents face a trade-off between the quantity (number) and the quality (education) of their children. Whereas the quantity of children prevailed during most of human history, parents began to prioritise child quality in the course of development. The increased investment in human capital spurred technological progress and economic growth. Ultimately, more child quality meant less quantity of children, reducing the

number of children, leading to lower fertility rates and causing the demographic transition. Therefore, the fertility transition was an important factor in the transition from the post-Malthusian era to the modern growth regime (see also Galor and Weil 2000). During our historical period, the demographic transition had already started in some regions whereas it was still to begin in others. Therefore, it is a relevant factor that we should include in our analysis. We use total fertility data provided by the famous Princeton European Fertility Project, which defines total fertility as “a measure of the fertility of all women in the population” (Coale and Treadway 1986, p. 154).

Moreover, marital status comes from the same source and is “the ratio of the number of births produced by married women in [...] a population to the number that would be produced if all women were married” (Coale and Treadway 1986, p. 154). In other words, this measure represents “the proportions married at each age” (Watkins 1986, p. 315) and can thus be used as a proxy for nuptiality. There have been important nuptiality differences in Europe in the past, as has most famously been put forward by Hajnal (1965). Hajnal pointed out that western Europe was characterised in the past by a specific and unique European Marriage Pattern (EMP). The EMP describes the fact that there were much higher average ages at marriage in western Europe than in eastern Europe (and the rest of the world). Thus, differences in the average age at marriage may also explain differences in economic development (e.g., de Moor and van Zanden 2010). For example, Foreman-Peck (2011) emphasises that this specific demographic pattern was an important force directly contributing to the development advantage of Western Europe by increasing innovation and productivity. Thus, we also control for nuptiality in our analysis.

In addition, population density is measured (in logarithmic terms) as the number of individuals per square kilometre. More generally, total population positively affects population growth and technological change in a very long-run perspective (e.g., Kremer

1993). Population density, as Klasen and Nestmann (2006, p. 623) point out, “generates the linkages, the infrastructure, the demand and the effective market size for technological innovations”. In this way, it may foster innovations and economic development in the long run. For this reason, population density has been a significant explanatory variable in empirical growth regressions in cross-country settings (e.g., Kelley and Schmidt 1995) and in some European countries (e.g., Ciccone 2002). Finally, its importance for technological progress and ultimately growth has been underlined in long-run growth models (e.g., Galor and Weil 2000). The data stem from Kirk (1946) in 1930. Population density has been derived for the other years from raster data provided by Klein Goldewijk *et al.* (2010) and Klein Goldewijk *et al.* (2011).³

The next variable is the share of the total population which is not dependent on agriculture. This share roughly proxies the regional economic development and industrialisation in 1930. Shares of agriculture or industry have been used in different historical publications where GDP per capita estimates are not available (e.g., Good 1994, Hatton and Williamson 1994, Becker and Woessmann 2009). Indeed, although we cannot show the relationship for historical GDP per capita estimates due to lack of data, Figure 2 shows that there is a relationship between this historical share and current GDP per capita. Some outliers are apparent, such as Luxembourg (LU00) but the general pattern clearly holds. For this reason, we argue that we can reliably proxy for historical economic development with this variable. Given the fact that we are interested in the correlation of historical variables with current economic development, it appears essential to control for the initial historical level of industrialisation. The data come from Kirk (1946).

³ To check whether these estimations are sufficiently reliable, we also correlated the derived data for 1930 with those calculated by Kirk (1946) in 1930. They are correlated to 91 %, allowing us to use them in our subsequent analyses.

In addition, infant mortality represents a variable related to health. According to Kalemlı-Ozcan (2002), low mortality may promote economic growth through different channels such as population growth and education. When parents face a high uncertainty about the survival of their children, they will demand a higher number of children. When the risk of child death is reduced, parents may increasingly replace child quantity by child quality. This decreases fertility and lowers human capital, leading to sustained long-run economic growth. Kirk (1946) provides this data.

Moreover, the capital region dummy has been introduced because capital regions have often specific characteristics due to their administrative functions. The dummy for the newer EU regions captures the fact that these countries joined the EU later on and have had different historical and economic experiences in the past, having mostly been part of the Communist bloc before the fall of the Soviet Union. More specifically, these regions come from the newest 12 EU members (Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Malta, Poland, Romania, Slovenia and Slovakia). For this reason, West Germany is also part of the ‘old’ member states, while (former Communist) East Germany is considered as part of the ‘new’ states even though it was already reunified with West Germany in 1990. Finally, there may be different inherent characteristics of countries (e.g., institutions) which may bias the results. Therefore, the inclusion of country dummies allows to control for these country fixed effects.

Most variables are available for 1930, which is why we focus in our analysis on this year. A reduced number of variables is also available for 1850, 1900 and 1960. These variables are literacy, fertility, marital status, population density and our two dummy variables. Descriptive statistics on all variables are shown in Table 1. We have up to more than 250 regions in our dataset at the different points in time. The regions that are covered may be different at each point, thus reducing the number of observations in the regressions.

In addition, we need to consider the question how a region is defined in this paper. Clearly, the regions in 1930 and at other points in time are often not the same as today, at least for a number of European countries.⁴ For this reason, the historical regions have been adapted to the NUTS classification of the European Union (see also related work by e.g., Hippe 2014). More precisely, we use NUTS 2 regions as our standard regional classification, which is also done in the relevant literature in regional economics (e.g., Badinger *et al.* 2004, Herwartz and Niebuhr 2011, Scherngell and Barber 2011).

Moreover, note that the availability of the data can be quite different at each time period. In particular, the Eurostat data for the current period refer only to countries of EU27, EFTA and some Candidate countries. For this reason, the corresponding regressions only consider these regions.

On the other hand, whereas the ABCC data for 1850 consider most of the European regions in the larger sense, the literacy data for 1900 and 1930 only refer to those countries where literacy was still measured. Still, many countries can be included in this study. In contrast, literacy in 1960 is only available for a reduced number of countries (see appendix). Therefore, the results for the data for 1960 are less comparable than for the other points in time. Still, they allow us to get some additional insights for the respective regions at the beginning of the second half of the 20th century.

4. Results

4.1 Relationship between patents per capita and GDP per capita today

Before analysing the long-run impact of human capital on innovation and economic growth, we consider the current relationship of the latter two dependent variables in our

⁴ For example, Spain and France have preserved almost the same regions and regional boundaries until today.

subsequent regressions. Figure 3 shows their relationship for 2008. The figure highlights a general positive relationship between current GDP per capita and patent applications per million inhabitants to the EPO in Europe.⁵ The ‘new’ member countries have typically a lower number of patents and GDP but they follow the basic pattern of the old member states, underlining the relevance of controlling for the new EU member states. The most important outliers are Inner London (UKI1) and Luxembourg (LU00) which had much higher GDP per capita levels than their relative number of patent applications would suggest. On the other hand, Germany’s core industrial zones in the greater region around Munich (Oberbayern, DE21) and Stuttgart (DE11) alongside Dutch Noord-Brabant (NL41) and Austrian Vorarlberg (AT34) apply most often. Finally, the lowest GDP per capita values have the regions in the two newest member states, i.e., Bulgaria and Romania.

4.2 Explaining regional patents per capita

In the next step, we use standard OLS regression models to dig deeper into the relationship between human capital and innovation on the one hand and between human capital and economic development on the other hand. More specifically, we regress current patents per capita (i.e., patent applications per million inhabitants, in 2008) on historical variables (in 1930). We use the year 2008 because it provides the highest number of observations.⁶ Note that we always include country dummies to control for country fixed effects. We report robust p-values to avoid problems related to heteroskedasticity. Nevertheless, all regions have the same weight, representing each an historical experience.

The results are highlighted in Table 2. In each case, literacy is a significant positive explanatory variable of current patents per million inhabitants at the 1 % level. In

⁵ Note that data are available for more regions in 2008 than in 2000.

⁶ Note that we will use an alternative range of years in subsequent robustness checks.

other words, when literacy increases by 1 %, patents per capita increase by 4.3 to 5.4 % – a sizeable effect. When all variables are included (column 1), population density is positively significant at the 10 % level, while newer EU regions have significantly lower patent applications (1 % level). This negative sign (in all cases except column 6) confirms the descriptive evidence shown in the figure above. When only literacy is considered, the dummy for capital regions turns significant (column 2), meaning that capital regions have a higher number of patents per capita than other regions. However, the coefficient is insignificant in all other cases.

These regression results show that literacy is the most significant historical explanatory variable for current patents per capita. However, how robust is this result? We propose several robustness checks. First, we perform a horse race, including only literacy and another explanatory in each regression to check whether our human capital indicator can survive the direct comparison with other potential explanatory variables (Table 3). These regressions confirm our previous results, indicating that literacy is the most important historical variable for explaining current patents per capita. Population density also appears to play a role, being significant (column 5). Capital regions (column 8) and newer EU regions (column 9) show also significantly higher and lower patent applications, respectively.

Second, a related question concerns multicollinearity. It is possible that some variables are highly correlated and this may cause biased estimates. In particular, fertility, marital status and infant mortality are potential candidates. We may consider this by excluding first one and then two of these variables from the regressions and check whether

the results are affected (not shown⁷). It turns out that literacy remains significant as before and the basic results hold.

Third, we may introduce an alternative measure for historical economic industrialisation: agricultural production per capita. It is highly correlated to the non-agricultural employment share (Figure 4). As non-agricultural employment is more closely conceptually related to economic development, this variable may potentially better represent historical productivity and innovative activities, particularly in those countries still dependent on agriculture. Some urbanised regions have a higher employment in non-agricultural sectors and constitute outliers in this respect. On the other hand, Danish regions were more productive than their employment share would indicate. In fact, Denmark was highly specialised in agricultural industry. We may test whether this alternative variable would change our results. In fact, in contrast to the non-agricultural employment share, historical agricultural production per capita is negatively significant at the 5 % level (Table 4). This new variable affects in particular population density, which is not significant anymore. As agricultural production per capita is a productivity indicator for this sector, it may mirror the typical productivity benefits of densely populated areas. Literacy, however, is still the most important driver of patents per capita. Its significance and coefficient remain largely stable, indicating the robustness of our previous results. We also re-run all previous regressions with agricultural productivity, but our results for literacy remain robust (not shown).

Fourth, until now we have used data for patent applications per capita in 2008. The reason for this choice is that this year has the highest number of observations. Still, one could question whether this year really represents the current period. It could potentially be a peculiar year, not representative for other years, and thus biasing our

⁷ All results not shown within the limits of this paper can be provided on request.

results. For this reason, we re-run all previous regressions employing an alternative time definition, using the average number of patent applications per capita between 2000 and 2010. However, note that data for the earlier years is not as available as for the more recent years, reducing our number of observations. Nevertheless, the use of this alternative dependent variable does not change our results (not shown).

Fifth, we claim that human capital is the most important historical indicator for current patent applications. We have demonstrated this by using historical data for around the year 1930. Similar to our reasoning above, one could argue that 1930 could be a special year which would not be representative of ‘the past’. In consequence, we check the robustness of our results using other years back in time. In particular, we use numeracy data for 1850 and literacy for 1900 and 1960. Our results should be broadly confirmed by these other years. While the data for literacy in 1900 is similar to those in 1930, remember that literacy is available for a reduced number of regions in 1960. Similarly, the coverage for numeracy in 1850 is different, and numeracy is a different human capital indicator. Therefore, we would expect the most similar results for literacy in 1900, while numeracy in 1850 and literacy in 1960 should confirm the broader picture. As mentioned above, the number of variables is importantly reduced in these alternative years. Therefore, we are only able use the following explanatory variables: human capital (numeracy or literacy), fertility, marital status, population density and the dummy variables (capital regions and newer EU regions). In consequence, this specification corresponds to columns 2 to 5 in our baseline regressions in Table 2. Note that we have these variables for the same reference years as literacy, while in the case of numeracy fertility and marital status are only available in 1870, i.e. 20 years later than numeracy and population density. While one can argue that this is a reasonable approximation for 1850, this approximation certainly results in an additional bias that we have to take into account in its analysis. The numeracy results

are, therefore, more tentative than those for literacy. The results are shown in Table 5. The human capital variable is in each case positive and highly significant, and most of the time significant at the 1 % level. In 1850 and 1900, population density is also positively significant, while fertility is highly negatively significant in our reduced sample for 1960. The coefficient is also relatively high for literacy in that year. The lower number of observations and thus the concentration on fewer countries may be an important reason for this. In addition, capital regions and newer EU regions are in several cases significant. Comparing these results to 1930, we see that they show the robustness of the human capital effect. The relevance of population density may have decreased over time, as its significance goes down as we come closer to the current time. On the other hand, comparing the literacy coefficient in our literacy regressions from 1900 to 1960, it appears that it is continuously increasing. The reduced sample in 1960 may have exaggerated this general tendency. The increasing coefficient may potentially show the increasing relevance of human capital over time for current patent applications. In any case, this last robustness check is in line with our baseline results and confirms the importance of historical human capital on current regional innovation patterns.

4.3 Explaining regional economic development

Let's now turn to explaining current regional economic development. We use regional GDP per capita (in PPS) in 2008 as our dependent variable and reproduce exactly the same strategy as for patents per capita.

Globally, the results are similar to those previously shown for innovation (see Table 6). Literacy is a highly significant explanatory variable of current GDP per capita (i.e., in 2008) at the 1 % significance level. A rise in literacy in 1930 by 1 % increases regional GDP per capita in 2008 by 0.83 to 1.05 % (depending on the specification), so

that there is an important influence of human capital. This result confirms the hypothesis that historical human capital is important for economic development in the long run. In the overall specification in column 1, marital status is also negatively significant, meaning that those regions where couples married on average earlier in 1930 have lower current GDP per capita in 2008. This appears to be in line with the assumption that early marriage may have negative consequences on economic development. Moreover, the capital regions have significantly higher and the newer EU regions significantly lower GDP per capita. The other explanatory variables do not have any significant effect. If we consider the other specifications in columns 2 to 5, we find that fertility may also negatively affect current GDP per capita at the 10 % level, but this only applies to column 3 and 4. This would mean that a higher fertility rate may have a negative effect on subsequent economic prosperity. This is in line with our expectations and the literature. However, in the other columns, this effect vanishes and the variable is insignificant.

In the next step, we proceed with the horse race between literacy and the other explanatory variables (Table 7). Literacy is always negatively significant at the 1 % level. In addition, most other variables are significant and show the expected signs. As before, the fertility coefficient is negative and significant. Marital status, competing with literacy, becomes insignificant. A higher population density in 1930 significantly increases GDP per capita, illustrating the potential positive effect of a dense population. In contrast, a rise in infant mortality has a negative significant effect (at the 10 % level), and the employment share in sectors other than agriculture is positive and significant. However, the previous results suggest that when we include all variables only the literacy effect survives (apart from the significance of the two dummy variables).

We further explore the issue of multicollinearity. One or two of the fertility, marital status and infant mortality variables are dropped in different specifications (not

shown). However, this does not affect any significance level or even sign in our baseline specification in Table 6 column 1. Literacy remains a stable significant explanatory variable.

Now, let us replace the employment share in sectors other than agriculture by agricultural productivity (Table 8). As its predecessor, it is not significant. Only marital status becomes insignificant. If we rerun the regressions with this new variable, we do not find any relevant changes in our results either.

In addition, we may replace the independent variable, GDP per capita in 2008, by the average GDP per capita between 2000 and 2010. Thanks to the database by Eurostat, we do not lose observations as in the case of patents. However, this alternative definition of our independent variable does not change our results and confirms once more the importance of human capital (not shown).

Finally, we consider other historical points in time to explain current GDP per capita (Table 9). Using numeracy data for 1850 and literacy data for 1900 and 1960, human capital appears as the only significant determinant in all specifications. The potential bias of including variables for 1870 in our numeracy specification applies similar to our patent regressions. In 1850 and 1900, marital status is significant and negative, similar to our results for 1930. Furthermore, population density is significant in both 1850 and 1900, while its significance vanishes in 1930 and 1960. Instead, fertility is negatively significant in 1960 in our reduced sample. Capital regions are positively and newer EU regions negatively significant in almost all specifications, confirming once again our results for 1930.

These results for human capital and other indicators at different historical points in time suggest that those regions that had a higher endowment in human capital in the past, that is even more than one hundred years ago, have higher GDP per capita levels

today than those regions which lagged behind. Moreover, capital regions are more prosperous than other regions.

To conclude, we find a positive and significant influence of historical human capital on current innovation and economic development. Human capital appears to be the most important factor contributing to today's innovation and economic development in our analysis. This suggests that human capital formation in Europe at the regional level is an important driver of economic development in the long run.

5. Conclusion

This paper has focused on the relationship between human capital, innovation and economic development in the European regions in a long-term perspective. There already exists a large literature on the effects of human capital on economic growth (e.g., Demeulemeester and Diebolt 2011) and regional human capital on economic development (e.g., Gennaioli *et al.* 2013). Globally, human capital is assessed to be crucial for regional economic development today. But is this a persisting effect? So far, there is (almost) no evidence for the regional level in most of Europe in the long run. Therefore, by using a large and new dataset we analyse the relationship between historical human capital and current economic indicators in the European regions.

We have employed different indicators of human capital, innovation and economic development. These proxies are literacy and numeracy for human capital, patent applications per million inhabitants for innovation and GDP per capita (in PPS) for economic development. Regions have been defined according to the NUTS classification system set up by the European Union to allow a maximum of comparability throughout time. Human capital is proxied by literacy in 1930. We add further control variables, such as fertility, nuptiality, infant mortality, population density, share of employment in non-

agricultural sectors, agricultural productivity and dummy variables for capital regions and the regions of the newer EU countries. To check the robustness of our results we provide a number of robustness checks. We include and exclude the different independent variables and provide an additional definition of each dependent variable. In addition to 1930, we alternatively consider numeracy (i.e., age heaping) in 1850 and literacy in 1900 and 1960. In all cases, we include country dummies to account for country fixed effects.

The results show that human capital is the most significant historical factor to explain current patent applications per capita and current GDP per capita. Literacy is highly significant in all proposed specifications. In addition, population density is positively significant in a number of specifications for patents per capita. Newer EU regions have generally lower patents per capita than the ‘old’ member states. Similarly, these regions have lower GDP per capita. Capital regions have generally also higher GDP per capita levels than other regions. Population density also often positively significantly affects current regional economic standards, while a low age at marriage has often a negative impact. Yet literacy appears to be the dominant factor. Independent of the point in time considered between 1850 and 1960, the results indicate that human capital is a significant determinant of current regional innovative and economic disparities.

Therefore, our analysis suggests that historical human capital formation is important to explain current economic prosperity in the European regions. For this reason, it appears crucial not to neglect long-term evolutions that have key implications for today’s economic development. For this reason, still more advanced research on long-run human capital formation in the European regions appears necessary to better understand economic development in the past, present and the future.

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.
- Acemoglu, D. and D. Autor (2012). What Does Human Capital Do? A Review of Golding and Katz's The Race between Education and Technology, *Journal of Economic Literature*, 50 (2): 426-463
- Acemoglu, D., Johnson, S. and J. A. Robinson (2005). Institutions as the Fundamental Cause of Long-Run Growth, in: Aghion, P. and S. Durlauf (eds.). *Handbook of Economic Growth*, Amsterdam: Elsevier, 385-472.
- Acs, Z., Anselin, L. and V. Varga (2002). Patents and innovation counts as measures of regional production of new knowledge, *Research Policy*, 31: 1069-1085
- Aghion, P. and P. Howitt (1992). A Model of Growth through Creative Destruction, *Econometrica*, 60: 323-351.
- Aghion, P., and P. Howitt (1998). *Endogenous Growth Theory*, MIT Press.
- Aghion, P., and P. Howitt (2006). Joseph Schumpeter Lecture. Appropriate growth policy: a unifying framework, *Journal of the European Economic Association*, 4 (2-3): 269-314.
- Alesina, A. and D. Rodrik (1994). Distributive Politics and Economic Growth, *Quarterly Journal of Economics*, 109: 465-490.
- Andersen, B. (2001). *Technological Change and the Evolution of Corporate Innovation. The Structure of Patenting, 1880-1990*, Cheltenham/Northampton: Edward Edgar
- Arrow, K. J. (1962). The Economic Implications of Learning by Doing, *Review of Economic Studies*, 29 (3): 155-173.

- 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
- Barro, R. J and G. S. Becker (1989). Fertility Choice in a Model of Economic Growth, *Econometrica*, 57 (2): 481-501
- Barro, R. J. (1991). Economic Growth in a Cross Section of Countries, *Quarterly Journal of Economics*, 106 (2): 407-443.
- Barro, R. J. and J.-W. Lee (1993). International Comparisons of Educational Attainment, *Journal of Monetary Economics*, 32: 363-394.
- Baten, J. and J. L. van Zanden (2008). Book production and the onset of modern economic growth, *Journal of Economic Growth*, 13 (3): 217-235
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Chicago: University of Chicago Press
- Becker, G. S. (1981). *A Treatise on the Family*, Cambridge: Harvard University Press
- Becker, G. S., Murphy, K. M. and R. Tamura (1990). Human Capital, Fertility, and Economic Growth, *Journal of Political Economy*, 98 (5-2): S12- S37
- Benhabib, J. and M. M. Spiegel (1994). The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country and Regional U.S. Data, *Journal of Monetary Economics*, 34 (2): 143-173
- Cantwell, J. (1989). *Technological Innovation and Multinational Corporations*, Oxford/Cambridge: Basil Blackwell
- Cappellen, A., Fagerberg, J. and B. Verspagen (1999). Lack of regional convergence, in: Fagerberg, J., Guerrieri, P. and B. Verspagen (eds.). *The Economic Challenge for Europe: Adapting to Innovation Based Growth*, Cheltenham: Edward Elgar

- Ciccone, A. (2002). Agglomeration Effects in Europe, *European Economic Review*, 46 (2): 213-227
- Ciccone, A. and E. Papaioannou (2009). Human Capital, the Structure of Production, and Growth, *Review of Economics and Statistics*, 91 (1): 66-82
- Cipolla, C. M. (1969). *Literacy and development in the West*, Baltimore: Penguin Books
- Coale, A. J. and R. Treadway (1986). A Summary of the Changing Distribution of Overall Fertility, Marital Fertility, and the Proportion Married in the Provinces of Europe, in: Coale, A. J. and S. C. Watkins (ed.). *The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project*, Princeton: Princeton University Press, 31-181
- Cohen, D. and M. Soto (2007). Growth and Human Capital: Good Data, Good Results, *Journal of Economic Growth*, 12: 51-76.
- Cohen, W. M. and D. A. Levinthal (1989). Innovation and Learning: The Two Faces of R&D, *Economic Journal*, 99: 569-596
- Comin, D., Easterly, W. and E. Gong (2010). Was the Wealth of Nations Determined in 1000 BC?, *American Economic Journal: Macroeconomics*, 2: 65-97
- 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
- Cuaresma, J. C., Doppelhofer, G. and M. Feldkircher (2012). The Determinants of Economic Growth in European Regions, *Regional Studies*, DOI:10.1080/00343404.2012.678824
- David, P. A. (1990). The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox, *American Economic Review Proceedings*, 80: 355-361

- De La Fuente, A. and R. Doménech (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make?, *Journal of the European Economic Association*, 4: 1-36
- De Moor, T. and J. L. van Zanden (2010). Girl power: the European marriage pattern and labour markets in the North Sea in the late medieval and early modern period, *Economic History Review*, 63 (1): 1-33
- Demeulemeester, J.-L. and C. Diebolt (2011). Education and Growth: What Links for Which Policy?, *Historical Social Research*, 36 (4): 323-346
- Diamond, J. (1997). *Guns, Germs and Steel: The Fates of Human Societies*, New York/London: W. W. Norton
- Diebolt, C. and K. Pellier (2009). Vers une nouvelle histoire économique des brevets, *Brussels Economic Review – Cahiers économiques de Bruxelles*, 52 (3/4): 204-214
- Diebolt, C. and K. Pellier (2012). 400 ans de protection par les brevets, *Revue économique*, 63 (3): 611-621
- Engerman, S. and K. L. Sokoloff (2000). Institutions, Factor Endowments, and Paths of Development in the New World, *Journal of Economic Perspectives*, 14: 217-232.
- Eurostat (2014). Database. Regional Statistics, online, extractions from data sources “pat_ep_rtot” and “nama_r_e2gdp”, last accessed 23 August 2014.
- Fagerberg, J., Verspagen, B. and M. Caniels (1997). Technology, growth and unemployment across European regions, *Regional Studies*, 31: 457-466
- Foreman-Peck, J. (2011). The Western European marriage pattern and economic development, *Explorations in Economic History*, 48: 292-309
- Galor, O. (2005). 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. (2012). The demographic transition: causes and consequences, *Cliometrica*, 6: 1-28.
- Galor, O. and D. N. Weil (1996). The Gender Gap, Fertility and Growth, *American Economic Review*, 86 (3): 374-387
- Galor, O. and D. N. Weil (2000). Population, Technology and Growth: From the Malthusian Regime to the Demographic Transition, *American Economic Review*, 90 (4): 806-828
- Galor, O. and O. Moav (2002). Natural Selection and the Origin of Economic Growth, *Quarterly Journal of Economics*, 117: 1133-1192.
- 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*, 1: 105-164.
- Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F. and A. Shleifer (2004). Do Institutions Cause Growth?, *Journal of Economic Growth*, 9: 271-303.
- Goldin, C. D. and L. F. Katz (2008). *The Race between Education and Technology*, Cambridge: Belknap Press.
- Good, D. F. (1994). The Economic Lag of Central and Eastern Europe: Income Estimates for the Habsburg Successor States, 1870-1910, *Journal of Economic History*, 54 (4): 869-891
- Griliches, Z. (1990). Patent Statistics as Economic Indicators, *Journal of Economic Literature*, 28: 1661-1707

- Hajnal, J. (1965). European Marriage Pattern in Historical Perspective, in: Glass, D. V. and D. E. C. Eversley (eds.). *Population in History*, Chicago: Aldine Publishing Company
- Hanushek, E. A. and L. Woessmann (2008). The Role of Cognitive Skills in Economic Development, *Journal of Economic Literature*, 46 (3): 607-668
- Hatton, T. J. and J. G. Williamson (1994). What Drove the Mass Migrations from Europe in the Late Nineteenth Century?, *Population and Development Review*, 20 (3): 533-559
- 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. (2014). Regional human capital formation in Europe in the long run, 1850-2010, Working Paper.
- Hippe, R. and J. Baten (2012). Regional Inequality in Human Capital Formation in Europe, 1790 – 1880, *Scandinavian Economic History Review*, 60 (3): 254-289
- Kalemli-Ozcan, S. (2002). Does the Mortality Decline Promote Economic Growth?, *Journal of Economic Growth*, 7: 411-439
- Kelley, A. and R. Schmidt (1995). Aggregate Population and Economic Growth Correlations: The Role of the Components of Demographic Changes, *Demography*, 32 (4): 543-555.
- Kirk, D. (1946). *Europe's Population in the Interwar Years*, Princeton: Princeton University Press
- Klasen, S. and T. Nestmann (2006). Population, population density and technological change, *Journal of Population Economics*, 19: 611-626

- Klein Goldewijk, K., A. Beusen, de Vos, M. and G. van Drecht (2011). The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years, *Global Ecology and Biogeography*, 20 (1): 73-86
- Klein Goldewijk, K., Beusen, A. and P. Janssen (2010). Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3.1, *The Holocene*, 20 (4): 565-573
- Kremer, M. (1993). Population growth and technological change: one million B.C. to 1990, *Quarterly Journal of Economics*, 108 (3): 681-716
- Krueger, A. B. and M. Lindahl (2001). Education for Growth: Why and For Whom?, *Journal of Economic Literature*, 39: 1101-1136
- Lucas, R. (1988). On the Mechanics of Economic Development, *Journal of Monetary Economics*, 22 (1): 3-42
- Madsen, J. B. (2008). Semi-endogenous versus Schumpeterian growth models: testing the knowledge production function using international data, *Journal of Economic Growth*, 13: 1-26
- Madsen, J. B. (2010). The anatomy of growth in the OECD since 1870, *Journal of Monetary Economics*, 57: 753-767
- Mankiw, N. G., Romer, D. and D. N. Weil (1992). A Contribution to the Empirics of Growth, *Quarterly Journal of Economics*, 107 (2): 408-437
- 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
- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66: 281-302

- Moser, P. (2005). How do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs, *American Economic Review*, 95 (4): 1214-1236
- Nelson, R. R. and E. S. Phelps (1966). Investment in humans, technological diffusion, and economic growth, *American Economic Association Papers and Proceedings*, 56 (1-2): 69-75
- North, D. C. (1981). *Structure and Change*, New York: W.W. Norton Company Inc.
- Persson, T. and G. Tabellini (1994). Is Inequality Harmful for Growth?, *American Economic Review*, 84: 600-621.
- Pritchett, L. (2001). Where has All the Education Gone?, *World Bank Economic Review*, 15: 367-391
- Rodríguez-Pose, A. and R. Crescenzi (2008). Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe, *Regional studies*, 42 (1): 51-67
- Romer, P. M. (1986). Increasing Returns and Long-run Growth, *Journal of Political Economy*, 94 (5): 1002-1037
- Romer, P. M. (1990). Endogenous Technological Change, *Journal of Political Economy*, 99 (5): 71-102
- Schultz, T. W. (1961). Investment in Human Capital, *American Economic Review*, 51: 1-17
- Schultz, T. W. (1975). The Value of the Ability to Deal with Disequilibria, *Journal of Economic Literature*, 13 (3): 827-846
- Sianesi, B. and J. Van Reenen (2003). The Returns to Education: Macroeconomics, *Journal of Economic Surveys*, 17 (2): 157-200
- Solow, R. M. (1956). A contribution to the theory of economic growth, *Quarterly Journal of Economics*, 70 (1): 69-94

- Statistisches Bureau (1862). Schweizerische Statistik, Bevölkerung. Eidgenössische Volkszählung vom 10. December 1860. I. Lieferung, Statistisches Bureau des eidg. Departement des Innern, Bern: Walder & Schiller
- Sterlacchini, A. (2008). R&D, higher education and regional growth: Uneven linkages among European regions, *Research Policy*, 37 (6): 1096-1107
- Von Tunzelmann, G. N. (2000). Technology Generation, Technology Use and Economic Growth, *European Review of Economic History*, 4: 121-146
- Watkins, S. C. (1986). Regional Patterns of Nuptiality in Western Europe, 1870-1960, in: Coale, A. J. and S. C. Watkins (ed.). *The Decline of Fertility in Europe: the Revised Proceedings of a Conference on the Princeton European Fertility Project*, Princeton: Princeton University Press, 315-336

Table 1 **Descriptive statistics**

Variable	obs.	mean	sd	min	max
ABCC 1850	265	0.94	0.07	0.65	1.00
Total fertility 1870	265	0.40	0.09	0.23	0.65
Marital status 1870	265	0.54	0.12	0.28	0.81
ln(Pop. density 1850)	265	3.79	1.05	-0.41	6.39
Literacy 1900	192	0.57	0.29	0.13	1.00
Total fertility 1900	192	0.39	0.12	0.20	0.68
Marital status 1900	192	0.58	0.12	0.31	0.81
ln(Pop. density 1900)	192	4.05	0.93	0.74	5.99
Literacy 1930	192	0.74	0.20	0.22	1.00
Total fertility 1930	192	0.30	0.11	0.05	0.54
Marital status 1930	192	0.58	0.09	0.32	0.80
ln(Pop. density 1930)	192	4.16	0.98	0.67	8.82
Infant mortality 1930	192	0.13	0.05	0.04	0.30
Not dep. on agr. 1930	192	0.46	0.24	0.06	0.99
Literacy 1960	146	0.83	0.11	0.59	0.99
Total fertility 1960	146	0.22	0.05	0.12	0.50
Marital status 1960	146	0.61	0.08	0.48	0.82
ln(Pop. density 1960)	146	4.10	0.96	0.71	6.64
ln(GDP/c 2008)	256	10.04	0.36	8.88	11.31
ln(Patents/c 2008)	256	3.62	1.66	-1.59	6.26
Higher edu. attain. 2008	256	0.72	0.15	0.18	0.97
Capital	256	0.08	0.27	0.00	1.00
Newer EU regions	256	0.22	0.42	0.00	1.00

Table 2 Regional patent applications per capita in 2008

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(Patents/c 2008)					
Literacy 1930	5.42*** (0.001)	4.33*** (0.000)	4.45*** (0.000)	4.57*** (0.000)	4.66*** (0.000)	4.59*** (0.000)
Total fertility 1930	0.94 (0.717)		0.20 (0.912)	0.25 (0.894)	0.77 (0.678)	1.46 (0.526)
Marital status 1930	1.01 (0.645)			0.83 (0.684)	1.09 (0.563)	1.00 (0.582)
ln(Pop. density 1930)	0.32* (0.092)				0.20 (0.146)	0.21 (0.138)
Infant mortality 1930	1.29 (0.802)					-3.14 (0.531)
Not dep. on agr. 1930	-1.46 (0.126)					
Capital	0.21 (0.513)	0.49** (0.026)	0.32 (0.161)	0.32 (0.163)	0.05 (0.861)	0.07 (0.812)
Newer EU regions	-2.19*** (0.003)	-2.24*** (0.000)	-2.21*** (0.000)	-2.14*** (0.000)	-0.72*** (0.007)	-0.55 (0.116)
Constant	-2.61 (0.297)	0.80 (0.306)	0.63 (0.587)	0.04 (0.986)	-1.86 (0.314)	-1.76 (0.326)
Observations	129	157	145	145	144	144
R-squared	0.87	0.84	0.85	0.85	0.85	0.86

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Patents/c refers to patent applications to the EPO per million inhabitants. Country fixed effects included.

Table 3 Horse race between literacy and other variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln(Patents/c 2008)								
Literacy 1930	5.42*** (0.001)	4.56*** (0.000)	4.48*** (0.000)	4.66*** (0.000)	4.36*** (0.000)	4.30*** (0.000)	4.21*** (0.000)	4.33*** (0.000)	4.56*** (0.000)
Total fertility 1930	0.94 (0.717)		-0.25 (0.888)						
Marital status 1930	1.01 (0.645)			0.81 (0.689)					
ln(Pop. density 1930)	0.32* (0.092)				0.21** (0.025)				
Infant mortality 1930	1.29 (0.802)					-3.69 (0.348)			
Not dep. on agr. 1930	-1.46 (0.126)						0.01 (0.986)		
Capital	0.21 (0.513)							0.49** (0.026)	
Newer EU regions	-2.19*** (0.003)								-2.16*** (0.000)
Constant	-2.61 (0.297)	0.57 (0.453)	0.69 (0.554)	0.01 (0.994)	-2.23*** (0.000)	1.15 (0.228)	0.91 (0.345)	0.80 (0.306)	0.57 (0.453)
Observations	129	157	145	145	156	157	141	157	157
R-squared	0.87	0.83	0.85	0.85	0.84	0.83	0.85	0.84	0.83

Robust p-values in parentheses. Significance levels are as follows: *** p<0.01, ** p<0.05, * p<0.1

Table 4 Including agricultural employment or agricultural productivity

	(1)	(2)
	ln(Patents/c 2008)	
Literacy 1930	5.42*** (0.001)	5.47*** (0.000)
Total fertility 1930	0.94 (0.717)	0.59 (0.803)
Marital status 1930	1.01 (0.645)	2.37 (0.233)
ln(Pop. density 1930)	0.32* (0.092)	0.16 (0.287)
Infant mortality 1930	1.29 (0.802)	-2.34 (0.629)
Not dep. on agr. 1930	-1.46 (0.126)	
Capital	0.21 (0.513)	-0.04 (0.868)
Newer EU regions	-2.19*** (0.003)	-1.32*** (0.000)
ln(Agr. prod./c 1930)		-0.57** (0.025)
Constant	-2.61 (0.297)	-3.48 (0.153)
Observations	129	143
R-squared	0.87	0.86

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Patents/c refers to patent applications to the EPO per million inhabitants. Country fixed effects included.

Table 5 Other points in time (1850, 1900, 1960) and patents

	(1)	(2)	(3)
		ln(Patents/c 2008)	
ABCC 1850	4.43** (0.013)		
Total fertility 1870	-0.25 (0.883)		
Marital status 1870	-1.13 (0.315)		
ln(Pop. density 1850)	0.33*** (0.000)		
Capital	0.35 (0.137)	0.16 (0.566)	-0.01 (0.968)
Newer EU regions	-1.18*** (0.000)	-1.71*** (0.000)	-0.71** (0.030)
Literacy 1900		3.38*** (0.000)	
Total fertility 1900		-2.33 (0.220)	
Marital status 1900		0.60 (0.669)	
ln(Pop. density 1900)		0.24** (0.014)	
Literacy 1960			10.25*** (0.000)
Total fertility 1960			-7.04*** (0.003)
Marital status 1960			1.40 (0.594)
ln(Pop. density 1960)			0.18 (0.212)
Constant	-0.88 (0.589)	0.43 (0.652)	-4.77** (0.044)
Observations	198	143	88
R-squared	0.82	0.82	0.82

Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Patents/c refers to patent applications to the EPO per million inhabitants. Country fixed effects included.

Table 6 Regional GDP per capita in 2008

	(1)	(2)	(3)	(4)	(5)	(6)
	ln(GDP/c 2008)					
Literacy 1930	0.87*** (0.000)	1.05*** (0.000)	0.89*** (0.000)	0.84*** (0.000)	0.85*** (0.000)	0.83*** (0.000)
Total fertility 1930	-0.18 (0.661)		-0.55* (0.054)	-0.55* (0.051)	-0.47 (0.115)	-0.34 (0.344)
Marital status 1930	-0.57** (0.045)			-0.41 (0.105)	-0.37 (0.138)	-0.39 (0.131)
ln(Pop. density 1930)	0.03 (0.280)				0.03 (0.192)	0.04 (0.176)
Infant mortality 1930	-0.66 (0.406)					-0.59 (0.427)
Not dep. on agr. 1930	0.11 (0.455)					
Capital	0.29*** (0.000)	0.42*** (0.000)	0.31*** (0.000)	0.31*** (0.000)	0.27*** (0.001)	0.27*** (0.001)
Newer EU regions	-1.42*** (0.000)	-1.13*** (0.000)	-0.17** (0.013)	-0.12 (0.125)	-1.31*** (0.000)	-1.30*** (0.000)
Constant	10.47*** (0.000)	10.10*** (0.000)	9.40*** (0.000)	9.60*** (0.000)	10.44*** (0.000)	10.49*** (0.000)
Observations	134	163	151	151	150	150
R-squared	0.91	0.87	0.89	0.90	0.90	0.90

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

Table 7 Horse race between literacy and other variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ln(GDP/c 2008)								
Literacy 1930	0.87*** (0.000)	1.26*** (0.000)	0.93*** (0.000)	1.15*** (0.000)	1.11*** (0.000)	1.17*** (0.000)	0.95*** (0.000)	1.05*** (0.000)	1.26*** (0.000)
Total fertility 1930	-0.18 (0.661)		-0.96*** (0.006)						
Marital status 1930	-0.57** (0.045)			-0.42 (0.181)					
ln(Pop. density 1930)	0.03 (0.280)				0.13*** (0.000)				
Infant mortality 1930	-0.66 (0.406)					-1.30* (0.089)			
Not dep. on agr. 1930	0.11 (0.455)						0.58*** (0.001)		
Capital	0.29*** (0.000)							0.42*** (0.000)	
Newer EU regions	-1.42*** (0.000)								-1.06*** (0.000)
Constant	10.47*** (0.000)	9.90*** (0.000)	9.32*** (0.000)	9.12*** (0.000)	8.47*** (0.000)	10.10*** (0.000)	9.78*** (0.000)	10.10*** (0.000)	9.90*** (0.000)
Observations	134	163	151	151	162	163	146	163	163
R-squared	0.91	0.80	0.86	0.85	0.85	0.81	0.83	0.87	0.80

Robust p-values in parentheses. Significance levels are as follows: *** p<0.01, ** p<0.05, * p<0.1

Table 8 Including agricultural employment or agricultural productivity

	(1)	(2)
	ln(GDP/c 2008)	
Literacy 1930	0.87*** (0.000)	0.94*** (0.000)
Total fertility 1930	-0.18 (0.661)	-0.55 (0.115)
Marital status 1930	-0.57** (0.045)	-0.21 (0.470)
ln(Pop. density 1930)	0.03 (0.280)	0.02 (0.355)
Infant mortality 1930	-0.66 (0.406)	-0.36 (0.625)
Not dep. on agr. 1930	0.11 (0.455)	
Capital	0.29*** (0.000)	0.27*** (0.001)
Newer EU regions	-1.42*** (0.000)	-0.31*** (0.002)
ln(Agr. prod./c 1930)		-0.07 (0.205)
Constant	10.47*** (0.000)	9.22*** (0.000)
Observations	134	148
R-squared	0.91	0.90

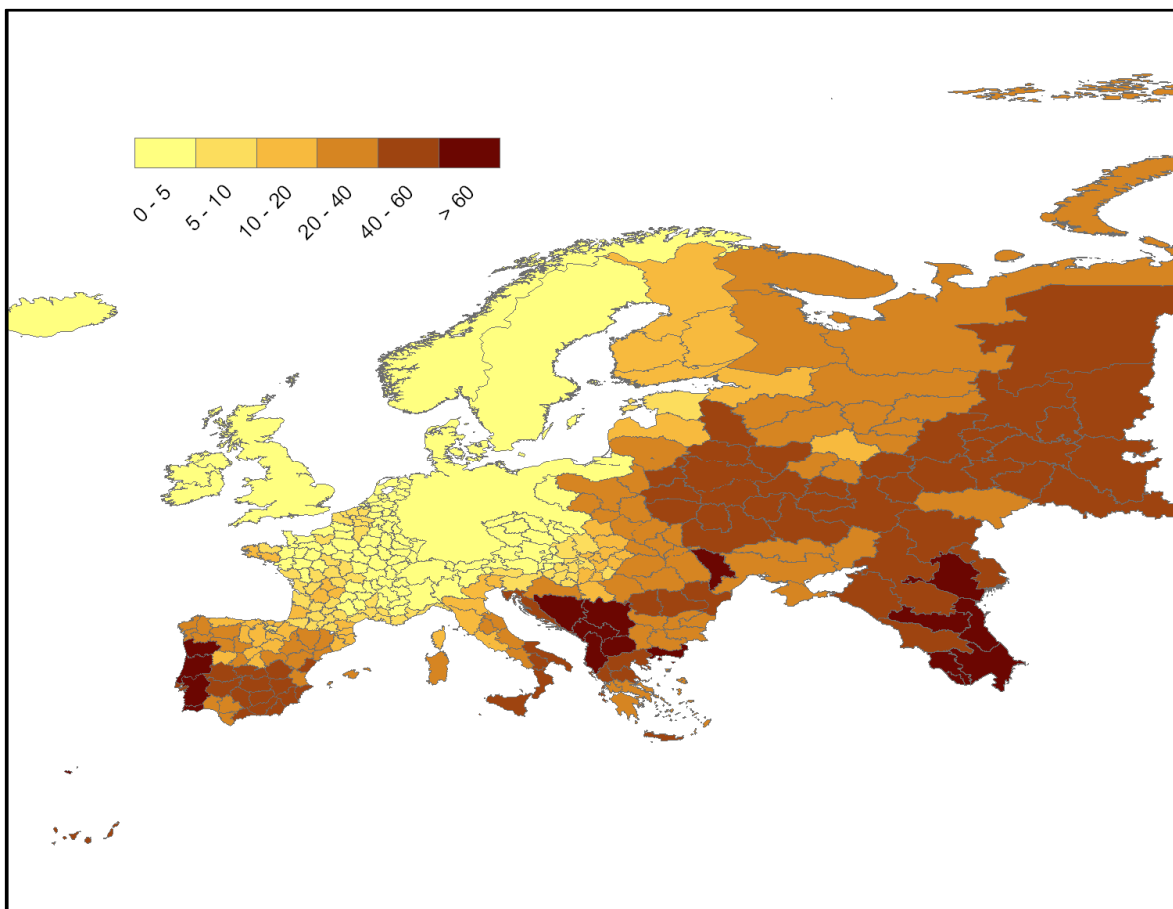
Note: ***, **, * indicate significance at the 1, 5 and 10 percent level. Robust p-values in parentheses. Country fixed effects included.

Table 9 Other points in time (1850, 1900, 1960) and GDP per capita

	(1)	(2)	(3)
		ln(GDP/c 2008)	
ABCC 1850	1.38*** (0.000)		
Total fertility 1870	0.09 (0.807)		
Marital status 1870	-0.53** (0.037)		
ln(Pop. density 1850)	0.10*** (0.000)		
Capital	0.34*** (0.000)	0.38*** (0.000)	0.36*** (0.000)
Newer EU regions	-0.37*** (0.000)	-0.49*** (0.000)	0.00 (0.982)
Literacy 1900		0.68*** (0.000)	
Total fertility 1900		0.27 (0.434)	
Marital status 1900		-0.61** (0.035)	
ln(Pop. density 1900)		0.05* (0.096)	
Literacy 1960			1.88*** (0.000)
Total fertility 1960			-1.72*** (0.002)
Marital status 1960			0.11 (0.837)
ln(Pop. density 1960)			0.02 (0.536)
Constant	8.66*** (0.000)	9.49*** (0.000)	8.60*** (0.000)
Observations	199	148	95
R-squared	0.79	0.83	0.88

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

Figure 1 **Illiteracy in 1930**



Source: Own calculations of illiteracy (in %), based on Kirk's (1946) data.

Figure 2 Non-agricultural employment, 1930 and GPD/c, 2008

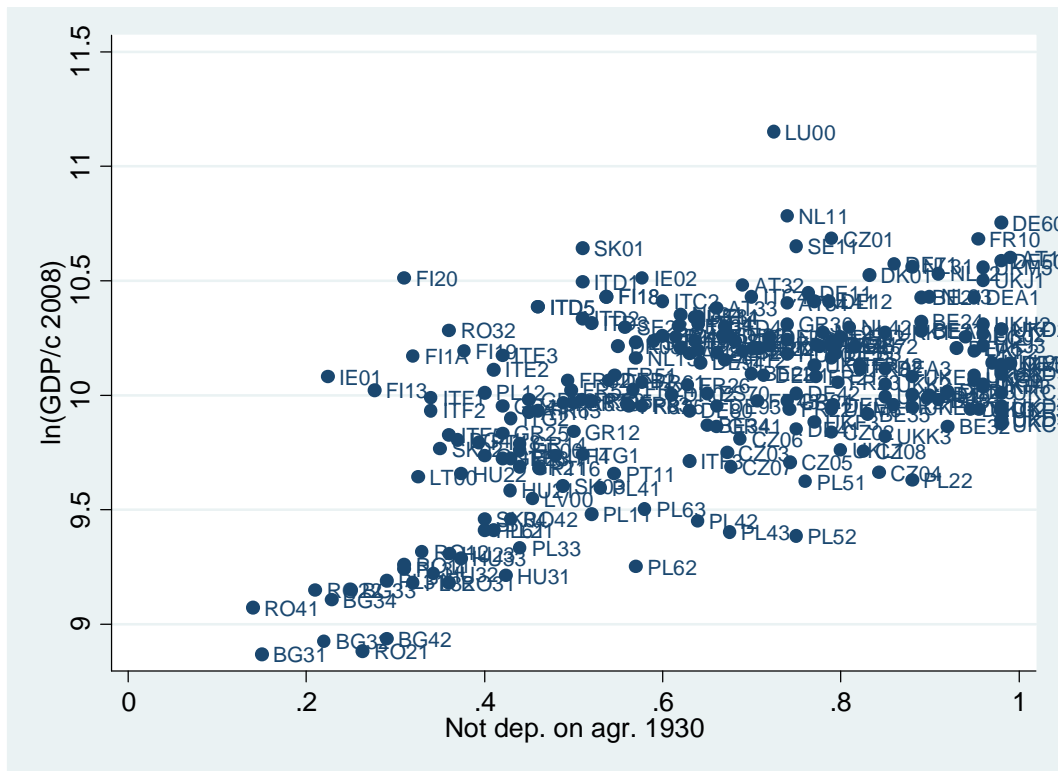
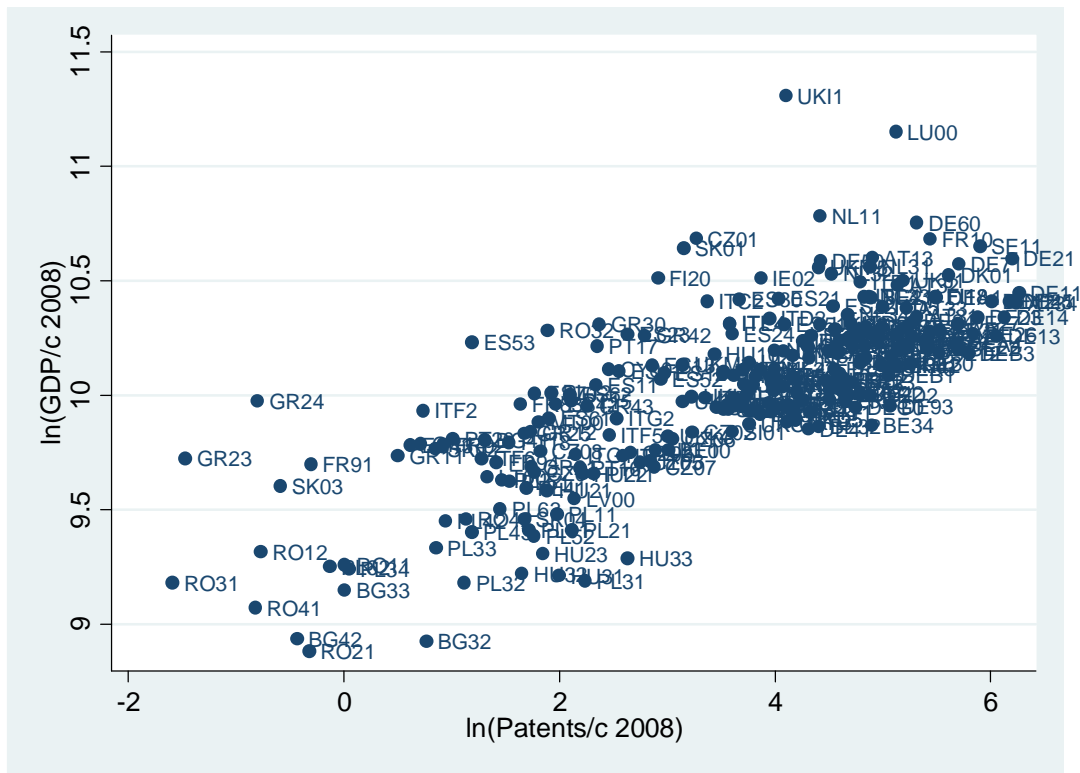


Figure 3 Regional per capita GDP and patent applications, 2008



Note: Patent applications per capita are defined as patent applications to the EPO per million of inhabitants.
Source: Data provided by Eurostat (2014).

Figure 4 Non-agricultural employment share and agricultural productivity



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

Regions of the following countries are included in the variables at each point in time:

year	country abbreviation
1850	AT, BE, BY, CH, DE, DK, EE, ES, FR, HR, HU, IE, IS, IT, LT, LU, LV, MD, ME, NL, NO, PL, PT, RO, RU, SK, SR, UA, UK
1900	AT, BE, BG, BY, ES, FR, GR, HR, HU, IE, IT, LT, LV, ME, PL, PT, RO, RU, SK, SR, UA, UK
1930	AT, BA, BE, BG, BY, EE, ES, FI, FR, GR, HR, HU, IE, IT, LT, LU, LV, MD, ME, MK, NL, PL, PT, RO, RU, SK, SR, UA
1960	BA, BG, EE, ES, GR, HU, IT, LT, LV, MD, ME, MK, PL, PT, RO, RU, SR
2008	AT, BE, BG, CZ, DE, DK, ES, FI, FR, GR, HU, IE, IT, NL, PL, PT, RO, SE, SI, SK, UK