Innovating to Invest: The Role of Basic Education

Francesco D'Acunto *

November 2014

Abstract

Despite the focus of entrepreneurial finance research on high-tech innovation, more than 75% of innovations are new processes and products in traditional manufacturing. I show that basic education is a key determinant of innovation in traditional industries. I document that manufacturers in European regions with 10% more high school graduates file 15% more patents, and invest 4% more in capital expenditures. To absorb spatially correlated unobservables, I construct Virtual Regions that only exploit the variation in basic education across nearby locations. To address the possibility of reverse causality, I establish that regional basic education persists for decades, and I use the quasi-exogenous diffusion of the printing press after 1450 to instrument for historical basic education. The results offer a human capital channel for innovation that feeds into the innovation-to-investment literature in finance.

JEL: D22, G32, I25, J24, N94, O31, R12.

---

*Haas School of Business, UC Berkeley. e-mail: francesco_dacunto@haas.berkeley.edu. For their precious and invaluable guidance I thank Stefano Della Vigna, Ulrike Malmendier, Ross Levine, Gustavo Manso, and Adair Morse. This research was conducted with restricted access to the EU-EFIGE/Bruegel-Unicredit data in Brussels (BE). The views expressed herein are those of the author, and do not necessarily reflect those of Bruegel or Unicredit. Tommaso Aquilante, Marco Antonielli, and Francesca Barbiero provided great help to access to the data. I thank Guido Tabellini for making his data on historical urbanization and the quality of European regional institutions publicly available, and Navin Ramankutty for his data on the quality of cultivable lands. For very helpful comments and discussions, I thank Johannes Buggle, Davide Cantoni, Barry Eichengreen, Laurent Fresard (discussant), Mariassunta Giannetti (discussant), Paola Giuliano, Luigi Guiso, Pat Kline, Martin Lettau, Petra Moser, Elias Papaioannou, Christine Parlour, Farzad Saidi, David Sraer, Annette Vissing-Jorgensen, Nico Voigtländer, Joachim Voth, Bas ter Weel, Michael Weber, Noam Yuchtman, and seminar participants at the 2014 NBER Summer Institute, the 2014 Western Finance Association, the 2014 SFS Finance Cavalcade, the 11th Annual Corporate Finance Conference at Olin, and UC Berkeley. Financial support from the White Foundation and the Clausen Center for International Business & Policy are gratefully acknowledged. All errors are my own.

"Our first priority is making America a magnet for new jobs and manufacturing. [...] Right now, countries like Germany focus on graduating their high school students with the equivalent of a technical degree from one of our community colleges."

President Obama, 2013 State of the Union Address

1 Introduction

Despite the limelight on high-tech innovation, 79% of the patents filed each year in the United States, more than 75% of those filed in Europe, and virtually all the unpatented innovation, are process and product improvements in traditional manufacturing industries (USPTO (2012), EuropeanCommission (2013)).[1] Traditional industries produce 54% of the American and 50% of the European manufacturing value added (BEA (2014)), but we know surprisingly little about the determinants and effects of their innovation.

In this paper, I study a neglected determinant of innovation. I show that the basic education of blue-collar workers increases the innovation activities of traditional manufacturing firms, and affects their investment and capital structure. Hence, firms can benefit from the formal education of blue-collar workers, and not only from their learning-by-doing within the firm. The results suggest that policies that discourage high school dropping out and increase the quality of high school education might be beneficial not only to workers, but to firms as well.

Patents in traditional industries, but not high-tech patents, increase with the amount of high school graduates (extensive margin of basic education), and with the years of schooling of inhabitants without a college degree (intensive margin of basic education) in European regions.[2] A 10% increase in the amount of high school graduates increases patents by 15%, and a 10% increase in years of schooling increases patents by 14%, after controlling for regional observables and limiting the variation within countries. The size of these associations is about 60% of the magnitude of the relationship between college education and patents in the sample.

[1] All manufacturing patents not produced in high-tech sectors based on the R&D-intensity-based definition of the *International Patent Classification* are traditional manufacturing innovation (WIPO, 2006). I enlist high-tech sectors in the Online Appendix.

[2] Focusing on European regions instead of US metropolitan statistical areas allows for the documentation of facts across countries that are exposed to different legal environments, institutions, and cultures. European regions are taken as a role model for the quality of their high school education in the policy debate on high school reform in the United States.

Kogan, Papanikolau, Seru, and Stoffman (2014) find that firms that innovate more have higher investment, and attract more capital.[3] Hence, basic education may affect firm-level investment and capital structure through innovation. In a unique data set on all the patented and unpatented innovation activities of 14,759 firms across 141 European regions, I confirm that higher regional basic education increases the product and process innovation of firms operating in the region. The effect is higher in firms with more basically educated employees, and is driven by traditional manufacturing firms. Consistent with Kogan, Papanikolau, Seru, and Stoffman (2014), basic education increases the capital expenditures and long-term debt of firms that innovate, but not of others.

The endogeneity of regional basic education to economic, institutional, and cultural dimensions poses a set of empirical hurdles. First, region-level policies and unobservables may determine both education and innovation. Second, unobservables such as local economic conditions, culture, and genetics, vary within regions and spill over across regions. Third, reverse causality may drive the results: regions with more innovation in traditional industries may attract better schools or individuals with higher incentives to stay in school, irrespective of an effect of schooling on innovation.

To address the issue of regional unobservables, I run the analysis at the level of European counties (NUTS 3 level), which allows averaging out any regional time-invariant effect, and I confirm the baseline results. The effects are larger in magnitude when excluding counties with at least one urban area above 200,000 inhabitants. Hence, consistent with the regional analysis, counties with a large scope for agglomeration economies do not drive the results.[4]

To address the issue of unobservables that vary at levels other than regions, I build on Michalopoulos (2012) and Michalopoulos and Papaioannou (2013) and develop a method, *Virtual Regions*, that allows me to control for unobservables flexibly within groups of

[3] A growing literature in finance studies the effect of external and internal financing on innovation inputs such as R&D expenditures, including the role of LBOs (Opler and Titman (1994); Lerner, Sorensen, and Stromberg (2011)), venture capital (Gompers (1995), Kaplan and Stromberg (2003), Kortum and Lerner (2000)), and angel investing (Kerr, Lerner, and Schoar (2014)).

[4] Commutes as short as 15 minutes allow spanning multiple counties in most regions; hence, the pool of blue-collar workers that firms can hire hardly coincides with the workers in the county where they operate.

neighboring counties. I impose a grid of 100 km by 100 km on the map of Europe.[5] The squares of land the grid creates are the Virtual Regions. The Virtual Regions' borders are arbitrary, and overcome the seldom-addressed issue of the endogeneity of regional borders. To construct the Virtual-Region data set, I sum the county-level variables for the counties that fall inside each Virtual Region.[6] For counties that are split across two or more Virtual Regions, I assign the county-level values to the Virtual Region that covers the largest part of the county.[7]

Because Virtual Regions do not fully overlap with regions, I can add regional fixed effects to the Virtual-Region specifications. To define the regional fixed effects, I assign each Virtual Region to the region that comprises its largest part. Thus, adding regional fixed effects means I only exploit the variation across the Virtual Regions that are mainly covered by the same underlying region. This procedure allows me to control for any spatially correlated unobservables common to neighboring Virtual Regions. The magnitude of the effect of basic education on patents remains similar to the baseline analysis: a 10% increase in the amount of high school graduates in a Virtual Region increases the Virtual-Region patents by 12%.

The configuration of the Virtual Regions and the shape of the grid are arbitrary. To test if the results are sensitive to the configuration, I repeat the analysis using four alternative shapes of the grid. For each shape, I move the grid in 27 positions, which repeatedly translate it: (i) 10 km east, (ii) 10 km north, (iii) 10 km north east. Alternative shapes and positions modify the composition of the Virtual Regions and the associated regional fixed effects. Across 108 alternative configurations, 106 estimated coefficients (98.1%) for the effect of basic education on patents are different from zero at the 5% level of significance. The mean of the coefficients is 12%, and the standard deviation is 2.5 percentage points. The estimated coefficients are similar across subsample analyses, such as excluding the Virtual Regions with the highest number of patents, the smallest Virtual

[5] I project the map of Europe with a cylindrical equal-area projection to avoid the distortions of geographical projections (see Dell (2009)).

[6] Figure A.2 in the Online Appendix plots the densities of some regional observables when reconstructed at the Virtual-Region level.

[7] The pieces of land in squares that do not cover the majority of any county, typically coastal areas, do not constitute a Virtual Region.

Regions,[8] the Virtual Regions with the highest population density, and each country at a time.

I then consider the issue of reverse causality between basic education and innovation. Ideally, one would need a source of exogenous variation in regional basic education at a time when manufacturing did not exist, and such variation should have persisted over time. I first establish that basic education persists in regions for decades, because literacy rates around 1880 predict the present-day amount of schooling for inhabitants with a high school degree or lower. This result is surprising, because European regions were differentially exposed to several institutional and economic shocks between the 19^{th} century and the present. Institutions and wealth are considered major determinants of human capital, but their abrupt changes across European regions have not modified significantly the pre-existing spatial distribution of basic education.

I then consider an epitomic shock to the incentives to acquire literacy in the past: the diffusion of the printing press after 1450 AD out of the city of Mainz, in current Germany (Dittmar (2011), Barbier (2006)). The printing press reduced substantially the cost of printed sources, and hence increased the incentives to acquire literacy. It was a quasi-proprietary technology, and it diffused concentrically out of the city of Mainz because of transportation costs. Being farther away from Mainz reduces regional literacy in 1880 and present-day basic education. It has no effect on college education, values, beliefs, or the quality of historical institutions.

The within-country distance from Mainz allows me to instrument for historical literacy to explain present-day innovation if it satisfies a demanding exclusion restriction: (i) the distance should not affect present-day innovation through historical channels different from historical basic education; and (ii) no unobservables should be systematically correlated with the distance and with innovation.[9] This exclusion restriction is untestable. I provide a set of results and placebo first-stage analyses that fail to document violations of the exclusion restriction. Historical literacy instrumented with the within-country distance from Mainz increases traditional manufacturing innovation

[8] I cannot split the largest Virtual Regions, because the maximal size of a Virtual Region is a full square, and 64% of the Virtual regions have full size.

[9] The exclusion restriction does *not* require that present-day basic education is the only channel through which historical basic education affects outcomes, because the distance from Mainz aims to instrument for historical literacy, and not for present-day basic education.

in regions and firms. The results do not change if I exclude Germany, Southern countries, or Eastern post-Communist countries.

The IV analysis allows rejection of the hypothesis that manufacturing innovation itself drives the whole relationship between basic education and current innovation and investment. As for the channels that may have transmitted the effect of historical literacy on innovation, I provide correlational evidence that the quality of historical and current regional institutions, college education, the dispersion of GDP within regions, or generalized trust, do not appear to have transmitted the effect of historical literacy on innovation. Present-day basic education appears to be a channel through which historical literacy affects innovation.

The main contribution of the paper is to shed light on a neglected driver of innovation in traditional industries, namely, the education of blue-collar workers. Moreover, the paper establishes that the quality of basic education is persistent in locations, and the economic and institutional shocks of the last decades did not significantly modify its spatial distribution. Traditional industries are neglected producers of innovation, and the results in this paper raise several questions for economics, finance, and strategy scholars, such as how the financing of high-tech innovation and innovation in traditional industries differ, how managers' and workforce characteristics interact to produce innovation, and the extent to which this type of innovation might boost economic development in countries without the resources for high-tech endeavors.

2 Related Literature

The paper belongs to four strands of literature in finance and economics.

A growing literature in finance studies the relationship between financing and innovation investment in firms, such as R&D investment. The focus has been on the effect of LBOs on innovation investment (Opler and Titman (1994); Lerner, Sorensen, and Stromberg (2011)), the role of venture capital in financing innovation (Gompers (1995), Kaplan and Stromberg (2003), Kortum and Lerner (2000)), and the effects of angel investing (Kerr, Lerner, and Schoar (2014)). In this paper, I build on Kogan, Papanikolau, Seru, and Stoffman (2014), who find that higher innovation brings firms to invest more and to attract more capital. I focus on the effect of innovation in traditional industries, which is incremental and requires little or no R&D investment, on the subsequent investment and financing of firms.

The paper contributes to the research that investigates the causes and consequences of innovation. The pioneering work of Schumpeter (1911) and Veblen (1904) shaped this area. Recent theoretical and empirical analyses include Manso (2011), Desmet and Rossi-Hansberg (2012), and Moretti and Wilson (forthcoming). Evidence exists on the role of input sharing and financing (e.g., Gompers and Lerner (2001)), peer effects (e.g., Lerner and Malmendier (2013)), wrongful discharge legislation (Acharya et al. (2013)), patenting laws (Moser (2005), Galetovic, Haber, and Levine (2014)), and agglomeration economies (e.g., Saxenian (1994), Chatterji, Glaeser, and Kerr (2013)), in the production of innovation. I contribute by studying the role of the formal education of workers on innovation in traditional industries, as opposed to specialized education and high-tech innovation, which is the focus of most previous literature. Formal education may complement the effect of learning-by-doing inside the firm (e.g., Levitt, List, and Syverson (2013)).

The paper also fits in the extensive body of micro- and macroeconomic research on the effects of human capital on firm-level and aggregate productivity and growth. Pioneering contributions are Mincer (1958), Becker (1962), Nelson and Phelps (1966), Easterlin (1981), Lucas (1988), and Barro (1991). Goldin and Katz (2008) dissect the relationship between mass education and technological advancement over time, and their effects on labor markets. Acemoglu and Autor (2012) discuss the interactions between skills and technologies and their implications for economic growth. Other contributions include Glaeser et al. (2004), Aghion et al. (2006), Ciccone and Papaioannou (2009), Becker et al. (2011), Schoellman (2012), Hanushek and Woessmann (2012), and Gennaioli et al. (2013).[10] I contribute by documenting the persistence of basic education in regions, and its effects on innovation and investment.

I also investigate the deep roots of present-day economic outcomes (Spolaore and Wacziarg (2013), Nunn (2013)). Acemoglu et al. (2001) document the long-run effects of colonial institutions on European colonies' growth. Putterman and Weil (2010) look at historical

[10]Whether human capital is a fundamental or proximate cause of growth is the subject of a long-term debate. The most recent discussion is in Acemoglu et al. (2014). The shock to the spatial distribution of historical basic education in section 7 is unrelated to the quality of historical regional institutions.

migratory patterns and growth. Dell (2010) studies the long-run effects of the Andean mining Mita on local outcomes, and Glaeser et al. (2014) study the long-run effects of carbon mines on the entrepreneurship of cities established nearby. Recently, historical arguments have helped explain present-day financial outcomes. In D'Acunto, Prokopczuk, and Weber (2014), past Jewish persecution reduces stock investments. In this paper, I use an historical shock to the distribution of literacy to study the effects of past basic education on innovation and investment.

3 Data

The unit of observation for the first part of the analysis is a European region. Regions are the NUTS 2 entities in the official administrative taxonomy of the European Union. The average region is a square with a side of 132 km (82 miles), and the average population is 1.8 million inhabitants.[11] The patenting and demographic regional data refer to 2005, which is the year I use for the cross-sectional analysis in the next section. All the results in the paper are similar if I use the regional data for any other year from 1998 to 2008, for which both patenting and demographic data are available at the regional level.

The regional data include geographic, demographic, and institutional characteristics. The main source for the geographic and demographic characteristics is the Eurostat Regional Database. Because the database does not include the average years of schooling by education levels, I measure them as the average years of schooling of the respondents to the World Value Survey Wave 9, covering the period from 1999-2004. This measure is not available for the regions of Austria, Belgium, Denmark, Greece, and Portugal. I measure the index of the quality of cultivable land at the regional level averaging the underlying 1 by 1 degree raster data of Ramankutty et al. (2002). I collect the first-wave results of the *EU Regional Competitiveness Index - RCI*, described in detail in Annoni and Kozovska (2010); I use the components on the quality of institutions and the quality of infrastructures. To obtain a regional measure of generalized trust, I average the individual level responses from the World Value Survey Wave 9 at the regional level.

[11] Regions generally coincide with the second-highest level of local government in each country, such as the *Regioni* in Italy. An exception is Germany, where NUTS 2 entities are the *Regierungsbezirke*, that is, subdivisions of the states.

The regional patent data are from the Patstat-Kites database, based on the European Patent Office database. I use the fractional count for the main analysis, but also establish the robustness of the results when using the integer count patent data and estimating negative binomial regressions to account for the count nature of the data.

I collect historical literacy data at the level of regions from Tabellini (2010) for the NUTS 2 regions included in the data set, and from other primary sources (national censuses) or secondary sources if not disaggregated at the NUTS 2 level in Tabellini (2010), which is the case for the regions of Germany and Spain. The literacy rate is the ratio of residents who could read and write in a European region around 1880. I can compute the measure for 228 regions for which national censuses existed in the second half of the 19^{th} century.

Panel A of Table 1 reports the regional statistics. Columns (1)-(3) refer to all available observations, and columns (4)-(6) refer only to the regions with non-missing historical data. On average, 173 patents are filed in each region, but the cross-regional variation is high. As for education, in the average region, 48% of inhabitants have high school degrees, whereas 23% of inhabitants have college degrees.[12] Figure A.4 in the Online Appendix plots the regional distribution of historical literacy. Positive spatial correlation of literacy exists across neighboring regions, which I will take into account in the empirical analysis. Regional literacy rates range from 15% in *Calabria* to 99% in the *Stockholm* region. The standard deviation of historical literacy is 26 percentage points, and large variation occurs within countries.

Firm-level data are from two sources: (i) Amadeus by Bureau van Dijk, which includes the financials of public and private companies incorporated in Europe;[13] and (ii) the EFIGE/Bruegel-Unicredit Dataset, based on a unique firm-level survey of 14,760 private and public firms in 141 European regions from seven countries (Austria, Italy, Spain, UK, Germany, France, Hungary). Firms were surveyed between 2008 and 2010. Altomonte and Aquilante (2012) describe the data set and the sampling procedures in detail. The firm-level responses include a large amount of soft information absent in the most

[12] In Figure A.1 of the Online Appendix, I plot the densities of regional variables across four countries, because I only use within-country variation in the analyses. Panel A of Figure A.4 plots the regional distribution of patents per capita.

[13] European private and public firms enter the Amadeus database if they fulfill at least one of three criteria: (i) operating revenues of 10 million euros or higher, (ii) total assets of 20 million euros or higher, and (iii) more than 100 employees. The sample does not include smaller firms.

common data sets used in corporate finance. The responses are organized around several areas that include the firm's ownership structure, workforce characteristics, innovation activities, foreign operations, and financing, among others. I create three dummies that equal 1 if a firm declares it engaged in any product, process, or both product and process innovations in the year prior to the interview. The definitions of product and process innovations are from the *Oslo Manual* (OECD, 2005): a product innovation is "the introduction of a good or service that is new or significantly improved with respect to its characteristics or intended uses." A process innovation is "the implementation of a new or significantly improved production or delivery method."

Panel B of Table 1 describes the financials, innovation, and firm-level financial constraints for all the firms with available data in both databases. Firms in the sample have on average 15 million euros in assets. Firm-level investment is capital expenditures (property, plant, and equipment) normalized by previous end-of-year assets. Tangibility - the ratio of tangible assets to total assets - is on average 0.26.[14] Firms have on average 117 employees. About 70% of the firms are controlled by a family or an entrepreneur. In the sample, 49% of firms innovated their products, and 32% innovated their processes. Seventeen percent of the surveyed firms were subject to financial constraints, either because their loan applications were rejected, or they had not applied because they expected a rejection, or they were only able to obtain a fraction of the capital they needed. Figure A.3 of the Online Appendix plots the densities of the most relevant firm-level variables across four countries.

4 Basic Education and Innovation in Regions

In this section, I document that basic education correlates positively with regional innovation in traditional manufacturing.

I first consider the extensive margin of basic education, that is, the amount of regional inhabitants with a high school degree. The following is my estimating equation[15]:

[14] I do not compute Tobin's Q or profitability measures based on market value, because more than 85% of the firms in the sample are private.

[15] I discuss the robustness of the results to using alternative specifications below, namely negative binomial regressions to account for the over-dispersed count nature of the patent data, and OLS regressions of the patents per capita on the ratios of regional inhabitants with a high school degree and a college degree.

$$Ln(Patents)_{r,c} = \alpha + \beta Ln(HighSchool)_{r,c} + \gamma Ln(College)_{r,c} + \theta Ln(NoDegree)_{r,c} + Ln(X)'_{r,c}\delta + \eta_c + \epsilon_{r,c}. \quad (1)$$

where $Ln(Patents)_{r,c}$ are the log of patents in region r, country c; $Ln(HighSchool)_{r,c}$, $Ln(College)_{r,c}$, and $Ln(NoDegree)_{r,c}$ are the log of individuals with high school, college, or no degrees living in the region; X are geographic, demographic, and economic regional controls; and η_c is a set of country fixed effects. The log-log specification is implied by the following Cobb-Douglas production function, which may describe the amount of patents produced in region r, country c:

$$Patents_{rc} = (\underbrace{HighSchool_{r,c}{}^{\beta}}_{\substack{\text{Inhabitants with}\\ \text{High School degree}}} * \underbrace{College_{r,c}{}^{\gamma}}_{\substack{\text{Inhabitants with}\\ \text{College degree}}} * \underbrace{NoDegree_{r,c}{}^{\theta}}_{\substack{\text{Inhabitants with}\\ \text{No Degree}}}) * X_{r,c}{}^{1-\beta-\gamma-\theta} * e^{\eta_c * u_{r,c}}. \quad (2)$$

In all specifications, standard errors are corrected to allow for correlation of unknown form within groups of regions (NUTS 1 level), as defined by Eurostat. A group of regions normally includes four or five neighboring regions. The standard errors are similar if I cluster them at the country level and correct for the low number of clusters following Cameron and Miller (2011), or if I allow for linearly-decaying spatial correlation in longitude and latitude following Conley (1999).[16]

I report the results for estimating Equation 1 in Panel A of Table 2. Within countries, a 10% increase in the amount of high school graduates is associated with a 14.6% increase in the amount of patents filed in the region (column (1)). The estimated association is about 60% as large as the association of patents with regional college graduates. Column (2) adds two controls - the average generalized trust from the World Value Survey, and the index of regional competitiveness - not observed for all regions, and the estimated association of high school graduates with patents is similar in magnitude. In columns (3)-(6), I find the association is driven by patents in traditional industries, whereas it is indifferent from zero for high-tech patents.[17] Figure 1 depicts this pattern. College educated inhabitants are important for high-tech patents, whereas their effect on patents

[16]I estimate the most conservative standard errors with a cutoff parameter of 2.5 degrees in longitude and latitude, and across all but one specification, the Conley standard errors are lower than those clustered at the level of groups of regions.

[17]The split between high-tech and mid-low-tech patents is not available for all regions.

in traditional industries is lower than the effect of high school graduates.

I move on to document the positive association between the intensive margin of basic education, that is, the years of schooling of those with high school degrees or below, and regional patents. I consider the following estimating equation:

$$Ln(Patents)_{r,c} = \alpha + \beta\theta Ln(s_{B,r,c}) + \beta Ln(BasicEducated)_{r,c} + \gamma\theta Ln(s_{C,r,c}) \\ + \gamma Ln(HigherEducated)_{r,c} + Ln(X)'_{r,c}\delta + \eta_c + \epsilon_{r,c}, \quad (3)$$

where $s_{B,r,c}$ is the average number of years of schooling for regional inhabitants with high school degrees or lower levels of education, $BasicEducated_{r,c}$ is the amount of regional inhabitants with high school or lower education, $s_{C,r,c}$ is the average number of years of schooling of those with at least some college education, and $CollegeEducated_{r,c}$ is the amount of regional inhabitants with more than high school education.
The average number of years of schooling is computed from the respondents to the World Value Survey, and is not available for all regions. Ten percent more years of schooling for those with basic education is associated with 13.7% more patents (column (1) of Panel B of Table 2). The association stays sizable and statistically significant for mid-low-tech patents (columns (5)-(6)).[18]

In Table 3, I show the regional results are robust to the exclusion of the regions that produce most patents, the regions with the highest GDP per capita, and the regions with the lowest GDP per capita. Moreover, I show the robustness of the results to estimating alternative specifications for the relationship between regional patents and basic education. In Panel D of Table 3, I estimate negative binomial regressions to account for the over-dispersed count nature of the patenting data. In Panel E, I regress the patents per capita in a region on the ratio of high school graduates, the ratio of college graduates, and the other regional controls.

To investigate which variation in the cross section of regions drives the results, in Panel A of Figure 2, I plot the coefficients estimated when sorting the regions in three groups based on the ratio of college-educated inhabitants. Regions with the lowest ratio of college-educated inhabitants drive the results. In Panel B of Figure 2, I sort regions by

[18] In column (3), the association between average years of schooling and high-tech patents is positive, but it loses statistical significance in column (4).

population density, which is relevant because (i) high-tech innovation clusters are mostly densely populated and urbanized areas, and (ii) the scope for agglomeration economies is larger in highly urbanized areas (e.g., see Carlino and Kerr (2014)). The least densely populated regions drive the associations.

5 Regional Unobservables: County-level Analysis

Endogeneity concerns plague the association between basic education and regional innovation. First, region-level dimensions such as policies and institutions, may drive both basic education and innovation. Second, several unobservable and non-measurable confounding variables, such as local economic shocks, attitudes toward risk, or ideology, vary within regions and spill over across neighboring regions. Third, reverse causality may explain the results.

To account for time-invariant regional unobservables, I exploit the variation across counties, or NUTS 3-level partitions in the official taxonomy of the European Union. The average county is a square with a side of 54 km (34 miles), and the average population is 311,000 inhabitants. On average, a region is composed of approximately six counties, although there is variation in the number of counties per region. Using the county as the level of observation allows for the absorption of regional fixed effects common to all the counties in a same region[19]:

$$
\begin{aligned}
Ln(Patents)_{k,r} = \alpha + \beta_k Ln(HighSchool)_{k,r} + \gamma Ln(College)_{k,r} + \theta Ln(NoDegree)_{k,r} \\
+ Ln(X)'_{k,r}\delta + \eta_r + \epsilon_{k,r}, \qquad (4)
\end{aligned}
$$

where $Ln(Patents)_{k,r}$ is the log of patents filed in county k of region r, and the other covariates are the same as in the regional analysis, but observed at the county level. Table 4 reports the estimated coefficient $\hat{\beta}_k$ from Equation 4. In column (1), a 10% increase in the number of high school graduates increases county-level patents by 15%; the estimated magnitude of the effect decreases to 11% once I control for regional unobservables in column (2). In columns (3) and (4), I exclude counties that host urban conglomerates with more than 500,000 and 200,000 inhabitants, because counties that

[19]The indices of institutional quality and of generalized trust are not available at the county level.

host large cities, have a high scope for agglomeration economies, or host industrial districts, may drive the results. Excluding counties that host cities with more than 200,000 inhabitants in column (4) reduces the sample by 46%. The size of the estimated coefficients is, if anything, higher once I exclude from the analysis the counties with large urban conglomerates. This result is consistent with the fact that the baseline association between basic education and patents at the regional level is driven by regions with low population density.

The analysis at the county level with regional fixed effects controls for any source of unobserved heterogeneity that varies at the regional level, such as regional policies or regional economic conditions. But the analysis is still subject to two shortcomings. First, commuting times as short as 15 minutes allow spanning several counties even in the largest regions (see Figure A.5 in the Online Appendix). The pool of blue-collar workers a firm can hire is likely dispersed across several counties, which suggests counties may not be the appropriate unit of observation for the analysis. Second, local economic shocks and unobservables that are likely to affect both the willingness to acquire education and the innovativeness of firms, such as local economic conditions, ideology, beliefs, or the average risk attitudes of the population, are likely to vary within regions and spill over across regions, because they were determined before regions existed.

6 Unobservables within and across Regions: Virtual Regions

To mitigate the concerns with the county-level analysis, I need a tool that allows me to flexibly control for local unobserved dimensions common to groups of counties within and across regional borders. I build on Michalopoulos (2012) and Michalopoulos and Papaioannou (2013) to construct Virtual Regions, that is, arbitrary aggregations of counties within and across regions.

To create the Virtual Regions, I project the map of Europe with a cylindrical equal-area projection to avoid the distortions of geographic projections (Dell (2009)), and I impose a grid of squares of 100 km by 100 km on the map. All the intersections of the squares with land are the Virtual Regions. The shape and configuration of the grid are arbitrary; hence, they cannot be endogenous to any economic, institutional, or political dimensions that have shaped the regional borders over time.[20]

I use the Virtual Regions as a new unit of observation. Figure 3 describes the construction of the Virtual-Region data set for the variable *Patents* across three Italian regions. Panel A reports the number of patents at the regional level for the three regions in the example. I sum up the county-level patents for the counties that enter each Virtual Region (Panel B). On average, three to four counties enter a Virtual Region. If a county is split between two Virtual Regions, I assign the county-level value to the Virtual Region that covers the majority of the county. This rule implies that small Virtual Regions, typically coastal areas, which cover no counties, are dropped from the analysis. I then sum up the regional portions of patents inside each Virtual Region (Panel C). Virtual Regions like A and B in Panel C of Figure 3 become the new units of observation.

There are more Virtual regions than regions.[21] The borders of the Virtual Regions do not coincide with those of the underlying regions. I index each Virtual Region by the underlying region $r*$ that covers the majority of its area. The index $r*$ emphasizes that Virtual Regions are not fully included in the underlying regions r. The following is the specification at the Virtual-Region level:

$$Ln(Patents)_{v,r*} = \alpha + \beta_v Ln(HighSchool)_{v,r*} + \gamma_v Ln(College)_{v,r*} + Ln(X')_{v,r*}\delta_v + \eta_{r*} + \epsilon_{v,r*}. \tag{5}$$

Equation 5 includes a full set of fixed effects for the underlying regions $r*$, η_{r*}.[22] These fixed effects are not collinear with the Virtual-Region variables, because the borders of the Virtual Regions do not fully overlap with those of the underlying regions. Adding the regional fixed effects allows demeaning all the variables by the average across the Virtual Regions indexed by the same $r*$:[23]

$$Ln(Patents)_{v,r*} - \overline{Ln(Patents)}_{r*} =$$

[20]I will show the results are insensitive to the shapes and configurations of the grid in the next section.

[21]Figure A.1 and Figure A.2 of the Online Appendix plot the distribution of variables across Virtual Regions.

[22]Of the 269 regions in the baseline analysis, 242 are associated with at least one Virtual Region.

[23]Below, the variable *College* is subsumed in the set of controls X.

$$\beta_v(Ln(HighSchool)_{v,r*} - \overline{Ln(HighSchool)}_{r*}) + (Ln(X')_{v,r*} - \overline{Ln(X')}_{r*})\delta_v + (\epsilon_{v,r*} - \overline{\epsilon}_{r*}), \quad (6)$$

where

$$\overline{Ln(HighSchool)}_{r*} = \frac{1}{N} \times \sum_{1}^{N} Ln(HighSchool)_{v,r*},$$

for the N neighboring Virtual regions that are indexed by the same $r*$. In the sample, the number of Virtual Regions indexed by the same $r*$ is one, two, or three. If only one Virtual Region is indexed by a $r*$, the information included in that square effectively drops from the analysis. When two or three Virtual Regions are indexed by the same $r*$, Equation 6 absorbs any time-invariant unobservables common to the two or three Virtual Regions. These areas are geographically close, because the largest Virtual Regions have a side of 100 km (62 miles); hence, unobservables such as the quality of institutions, values, beliefs, or risk attitudes, hardly vary across neighboring Virtual Regions. Virtual Regions allow me to interpret the estimated effect $\hat{\beta}_v$ in Equation 6 causally if such unobservables are the only threat to identification.

A county-level unobservable that is not spatially correlated across neighboring counties is not be absorbed in the Virtual-Region analysis. An example of such unobservables would be a set of policy interventions to increase the county-level basic education, and to increase the firm-level innovation, which only apply to individuals and firms residing in the county. These policies should be implemented in several counties across different Virtual Regions to drive the results. But note that the educational policies of European countries are mainly implemented at the national and regional levels. Policies like transfers or tax cuts to firms are mainly managed at the national level or through European Union funds that are distributed by regions.

Table 5 reports the estimated coefficient $\hat{\beta}_v$ in Equation 6 across four specifications. In columns (1)-(2), I restrict the variation within countries. This specification allows direct comparison between the estimated magnitude of the effect in the Virtual-Region data set and the regional data set. The comparison is important, because the Virtual-Region distributions are in general different from the regional distribution. The presence of 10% more high school graduates in a Virtual Region increases the patents filed in the Virtual Region by 11.6%. The effect is similar in size to the baseline regional analysis, in which the same increase in regional high school graduates is associated with 14.6% more patents filed in regions. In column (3), I add the regional fixed effects of Equation 6 to a specification without controls. The explained variation in the outcome is close to 1, and 10% more patents in Virtual Regions increase patents by 12.1%. In column (4), I add the covariates measured at the level of Virtual Regions to the specification, and the size of the estimated coefficient is similar. Adding the covariates allows me to address the concern that the procedure to create Virtual Regions may artificially produce systematic variation in the data. This concern does not drive the results, because adding the covariates does not reduce the magnitude of the estimated coefficient.

The last point can be formally tested using the Hausman-type decomposition discussed by Gelbach (2014), which allows derivation of a test statistic for the null that the estimated coefficients with and without covariates are the same. This null hypothesis is equivalent to the hypothesis that the covariate of interest is uncorrelated with the other covariates in the specification. The statistic is defined as follows:

$$t = \frac{\hat{\beta}_{baseline} - \hat{\beta}_{full}}{s.e.(\hat{\beta}_{baseline} - \hat{\beta}_{full})}, \tag{7}$$

where $\hat{\beta}_{full}$ and $\hat{\beta}_{baseline}$ are the estimated coefficients for the same OLS specification with or without the covariates of the model.[24] For the specification with country fixed effects, the null is rejected by construction, because the estimated standard errors in column (2) are higher than in column (1). For the specification with regional fixed effects, instead, we can compute the statistic in Equation 7, which is 0.35. Hence, after adding the regional fixed effects in the Virtual-Region specification, we cannot reject the null of no correlation between basic education and the other covariates at any plausible level of significance.

This test is relevant to exclude alternative explanations. For instance, cities or industrial districts in Virtual Regions cannot drive the effect; otherwise, adding population density

[24] Note that the statistic is derived based on the assumption that the model with covariates is correctly specified.

as a covariate would have decreased the coefficient on basic education.

In Table 6, I report a set of robustness checks and subsample analyses for the Virtual-Region specification.[25] In columns (1)-(2), I exclude the top 5% and 10% Virtual Regions by number of patents, whereas in columns (3)-(4), I exclude the bottom 5% and 10% Virtual Regions by area.[26] The size of the estimated coefficients is similar to the baseline analysis. In columns (5)-(6), I exclude the Virtual Regions with the highest population density. Consistent with the regional baseline analysis, the size of the estimated coefficient is higher when I drop from the analysis the most urbanized areas, which have a high scope for agglomeration economies.

The grid in the analysis so far was arbitrary. The optimal grid would combine neighboring counties with most unobservables in common. By definition, I cannot define the optimal grid. Absent a criterion to define the optimal grid, I investigate how sensitive the results are to the shape of the grid and to its position. I replicate the Virtual-Region analysis across several alternative configurations, which employ four different shapes and 27 alternative positions for each shape.

The shapes include: (i) 100 km by 100 km squares, like those in the analysis so far; (ii) isosceles triangles of side 100 km, obtained by dividing each square in half along its northwest to southeast diagonal; (iii) parallelograms of side 200 km and height 100 km, obtained by merging two triangles across different squares; and (iv) squares whose sides are the diagonals of the original squares. By construction, the four shapes produce Virtual Regions that generally differ from each other in their county composition.
As for the positions of the grid, I translate each grid in 27 alternative ways to explore alternative positions with non-overlapping grids: (i) 9 times east, in increments of 10 km; (ii) 9 times north, in increments of 10 km; and (iii) 9 times northeast, in increments of 10 km * $\sqrt{2}$.

Translated grids for each shape modify the composition of Virtual Regions as long as counties mainly covered by one square move to another square after the translation. Moreover, the grids produce alternative rules of regional indexing; hence, different groups

[25]Additional robustness results are in the Online Appendix.
[26]I cannot exclude the largest Virtual Regions, because the size of Virtual Regions is lower or equal to a full square, and 64% of the Virtual Regions in the sample have full size.

of Virtual Regions enter the same fixed effect.[27]

I run the specification in Equation 6 for each of the 108 shapes and positions of the grid, and hence obtain 108 estimated coefficients and standard errors. Table 7 reports the summary of the results.[28] Of 108 estimated coefficients, 106 (98.1%) are significantly different from zero below the 5% level of significance. The mean of the coefficients is 12%, and their standard deviation is 0.25. The mean of the estimated standard errors is 0.35. Because the mean of the estimated coefficients is close to 11.8%, the magnitude of the coefficient in the original Virtual-Region analysis (column (4) of Table 5), I conclude that the results of the Virtual-Region analysis are not sensible to the shape or position of the grid I use. Table 7 also reports the median and mean estimated coefficients separately for each shape of the grid. The within-shape means range from 11.7% to 12.4%.

7 Reverse Causality: Persistence of Basic Education and Historical Instrument

The analysis so far has shown that region-level unobservables, or unobservables common to counties within and across regions, do not drive the effect of basic education on local innovation. A remaining concern is that reverse causality drives the results: regions with more innovative firms may attract the best schools, or individuals that stay in school longer irrespective of an effect of schooling on innovation. To address this issue, I exploit an exogenous shock to the incentives to acquire basic education in the distant past - when manufacturing did not exist - and whose effects have persisted over time.

A. Persistence of Basic Education

First, I establish the persistence of basic education in regions. I show that a measure of basic education in the past, the historical regional literacy rate, predicts the present-day amount of basic education at the regional and individual levels, but not higher levels of education. For the regional level, I estimate the following specification:

$$Ln(HighSchool)_{r,c} = \alpha + \beta Literacy1880_{r,c} + X'_{r,c}\gamma + \eta_c + \epsilon_{r,c}, \quad (8)$$

[27]The Virtual Regions and fixed effects for which a 10 km translation changes none of the component counties will stay the same.

[28]In the Online Appendix, I report each coefficient estimated across the shapes and grid positions.

where $Ln(HighSchool)_{r,c}$ is the log of the regional population with a high school degree[29] but not higher education in region r and country c; $Literacy1880_{r,c}$ is the regional literacy rate around 1880; $X_{r,c}$ is a set of region-level geographic, historical, and demographic characteristics; and η_c are country fixed effects. In Table 8, a 10-percentage-point increase in literacy in 1880 is associated with a 2% increase in the number of current regional high school graduates (columns (1)-(2) of Panel A). Instead, historical literacy is uncorrelated with the current amount of college-educated individuals (columns (3)-(4)). One may be concerned that regions with higher past and current basic education attract more immigrants, but this concern is inconsistent with the results in columns (5)-(6) of Table 8.[30] The persistence of basic education from 1880 onward is surprising because European regions were differentially exposed to severe institutional and economic shocks from the end of the 19th century onwards, including two World Wars, the economic crisis of the 1930s, and the experience of totalitarian political regimes.

For the individual level, historical literacy rates in regions predict the years of schooling of current inhabitants with basic levels of education, but not of others. I consider the following:

$$YearsSchooling_{i,r,c} = \alpha + \beta Literacy1880_{r,c} + X'_{r,c}\gamma + D'_{i,r,c}\delta + \tau_{i,r,c} + \eta_c + \epsilon_{i,r,c}, \qquad (9)$$

where $YearsSchooling_{i,r,c}$ is the number of years of schooling for individual i in region r and country c, which take parts in the World Value Survey; $D_{i,r,c}$ is a set of respondent characteristics; and $\tau_{i,r,c}$ are five dummies for town size, which aim to eliminate the systematic variation across individuals living in rural or urban areas. Individual characteristics are available for about one half of the observations. In Panel B of Table 8, a 10-percentage-point increase in historical literacy is associated with 0.35 additional years of schooling in the full sample (column (1)). A major concern is that those who repeat one or more years of schooling before obtaining an academic degree drive the results. This concern is compelling because most countries impose age-based

[29]The results are similar if I use the share of regional inhabitants with a high school degree as the dependent variable.

[30]I focus on within-country immigrants, measured by Eurostat, because (i) out-of-country immigrants are often not registered and (ii) within-country immigrants are usually skilled workers attracted by job opportunities (Moretti, 2012), whereas the motivations of other immigrants may be different.

thresholds before citizens can legally withdraw from school.[31] In column (2), I exclude the variation in schooling across academic degrees,[32] and reassuringly, the effect of historical literacy on schooling disappears: those who stay in school longer because they repeat classes do not drive the persistence result. In columns (3)-(6), I find persistence holds only for individuals with basic education, and not for those with higher levels of education.

B. Instrument for Historical Basic Education

To address the possibility of reverse causality with the effect of basic education on innovation, one would need a source of exogenous variation in basic education at a time when manufacturing did not exist, and build on the persistence of basic education in regions. In this section, I describe the quasi-exogenous diffusion of the printing press after 1450 out of the city of Mainz, in current Germany Dittmar (2011), which is close to such as ideal exogenous shock.

The printing press was invented around 1450 by zu Gutenberg in Mainz (Pfalz), and it was a quasi-proprietary technology. The printing-press technology represented a positive shock to the incentives to acquire literacy for 15th-century European households, because it dramatically reduced the cost of printed sources. In modern Europe, delivering books from the cities where they were printed was as expensive as the monthly wage of a skilled craftsman (Dittmar, 2011). Printing presses started to spread around the city of Mainz through zu Gutenberg's own collaborators. The diffusion path was slow and concentric out of the city of Mainz because of transportation costs (Barbier, 2006). The treatment of having a printing press early across regions at a similar distance from Mainz can be assumed to be as good as randomly assigned. Figure 4 describes the pattern of diffusion of the printing press across European regions.

I detect negative unconditional correlation between the Euclidean distance of a region from Mainz and the literacy rate in the region in 1880 (Panel A of Figure 5). I discuss the within-country variation in regional distances from Mainz below. The distance from Mainz is unconditionally negatively correlated with the present-day ratio of inhabitants

[31]Strikingly, these regulations are often not enforced. OECD (2013) describes early school leaving across European countries, a phenomenon that varies substantially within countries.

[32]The terminal years of degree cycles vary across countries. I follow the WVS taxonomy for six international levels of education conducive to degrees across different European countries.

with high-school degrees, but it is unrelated to the present-day spatial distribution of college education (Panel B of Figure 5).

Whether education is a fundamental cause of economic growth is the subject of a long-term debate.[33] In this paper, I do not study GDP growth. Innovation in traditional industries likely contributes to regional growth, but it needs not be the most important driver of growth (e.g., Gennaioli et al. (2013)). At the same time, testing whether the exogenous variation in the diffusion of the printing press may have determined the quality of historical institutions, or other dimensions that drove growth. In Figure A.8 of the Online Appendix, I plot the unconditional correlation between the distance from Mainz and several historical, geographic, and current observables, and I do not detect patterns similar to those for literacy rates in 1880.

The distance from Mainz can be a valid instrument for historical literacy only if it does not affect current innovation through historical channels unrelated to historical literacy, and only if there are no unobservables that are correlated with both the distance from Mainz and current innovation. The exclusion restriction, which is quite demanding, cannot be formally tested. In Table 9, I provide a set of results that aim to inform on the plausibility that the exclusion restriction holds.

First, the minimal Euclidean distance of the centroid of a region from the city of Mainz is not associated with region-level and firm-level observables.[34] In Panel A of Table 9, no associations are economically or statistically different from zero, except for a negative association of the distance with the leverage of firms (columns (1) and (3)).

Second, the distance from Mainz may have determined historical economic growth, or other historical dimensions that favored current innovation but are no more observable. To investigate this possibility, I create a purged measure of distance, which consists of the residuals from an OLS regression of the minimal Euclidean distance from Mainz on a set of geographic and historical observables that may have affected the past economic conditions of regions: the latitude, area, quality of cultivable lands (Ramankutty et al.,

[33]Acemoglu et al. (2014) is the most recent contribution that summarizes the debate.

[34]Detecting no associations is not evidence that the exclusion restriction holds, because (i) by definition, I cannot test if any unobservables are correlated with the distance, and (ii) the imprecision of the estimated effects may drive the non-results.

2002), whether the region hosted any cities in the Hanseatic League, whether the region was Catholic after the Peace of Augsburg of 1555, and whether the region was in the Communist block after the Second World War. I then regress current region-level and firm-level observables on this purged measure of distance. In columns (2) and (4) of Panel A of Table 9, I find no significant associations of the purged distances with any of the current region- and firm-level observables.

Third, I propose two reduced-form specifications, in which the log of the distance from Mainz enters as a covariate together with other regional characteristics, and with the log of historical literacy (columns (2)-(3) of Panel B of Table 9).[35] The distance from Mainz is negatively associated with current regional innovation when it enters alone, but the estimated association drops in magnitude by about 30% and becomes statistically insignificant when the distance enters the same specification as historical literacy. This result is not easily compatible with an autonomous association of the distance from Mainz with present-day innovation.

Fourth, I run placebo first stages. I predict historical literacy with the distance of a region from cities other than Mainz, which the economic history literature has described as wealthy and proto-industrial in the 15^{th} and 16^{th} centuries (Prague, Amsterdam), or the cradle of nation states and national politics (London, Madrid). I also look at Florence, which was the cradle of Humanism after the 15^{th} century, and Aix-la-Chapelle (Aachen), where Holy Roman Emperors were enthroned as of the 15^{th} century, which is close to Mainz. Panel C of Table 9 reports the first-stage statistics for these placebo first stages, and it shows that none of the alternative distances seems to satisfy the relevance assumption for an instrument of historical literacy, although the magnitudes of first-stage test statistics increase with the correlation of each placebo distance with the distance from Mainz.

I run a two-stage least-squares analysis at the regional level, where historical literacy is instrumented with the distance from Mainz.[36] Panel A of Table 10 reports the outcomes of the second stage, where the dependent variable is the log of regional patents. In Panel B, I report the first-stage statistics: (i) the Cragg-Donald F-statistic, which is based on

[35]Column (1) of Panel of Table 9 coincides with column (2) of Panel A of Table A.1.

[36]In the Online Appendix, I report the results for regressing current innovation and firm-level outcomes on historical literacy.

i.i.d. standard errors, but is used to compute the critical values reported in Table 5.2. of Stock and Yogo (2005); (ii) the Kleibergen-Paap F-statistic, which is computed for correcting the standard errors for correlation of unknown form at the level of groups of regions; and (iii) the Angrist-Pischke chi-square statistic, which can be used for a rank test of the matrix of the reduced-form equation coefficients and the excluded instruments. In column (1) of Table 10, a one-standard-deviation increase of the instrumented log of literacy in 1880 increases the current log of regional patents by 0.57 standard deviations. The size of this estimated effect is about twice the size of the corresponding reduced-form specification (column (2) of Table A.1). The tests for weak identification and underidentification that use the Kleibergen-Paap F-statistic and the Angrist-Pischke chi-square statistic reject the null hypotheses based on the Stock-Yogo critical value for a worst-case IV bias size of 15% or lower. The results survive when excluding the regions of ex-Communist countries, of southern Europe, of Germany, or all three groups.

8 Basic Education and the Investment and Financing of Firms

In this section, I study the implications of the effect of basic education on innovation for the investment and financing of traditional manufacturing firms. Kogan, Papanikolau, Seru, and Stoffman (2014) find that firms that innovate more have higher investment, and attract more capital. Hence, basic education should affect firm-level investment and capital structure through innovation. I use a unique data set on the patented and unpatented innovation activities of 14,579 firms to document that manufacturers in European regions with higher basic education are more likely to innovate their processes and products than manufacturers in other regions. They also invest more in capital expenditures and raise more long-term debt as a ratio of total debt.

Several manufacturing innovations are never patented, because patenting requires financial and organizational resources. Observing *all* the innovation activities of firms, both patented and unpatented, is therefore crucial to studying innovation in traditional manufacturing. I use the unique EU-EFIGE/Bruegel database to estimate the following probit specification:

$$Pr(Innovation = 1)_{frc} = \Phi(\alpha + \beta Ln(HighSchool)_{rc} + X'_{rc}\gamma + F'_{frc}\delta + \eta_c + \eta_a + \eta_s + \eta_l). \quad (10)$$

Across three specifications, $Innovation$ is a dummy equal to 1 if the firm declares it engages in product, process, or both types of innovation. X_{rc} and F_{frc} are region- and firm-level covariates, and η_a, η_s, and η_l are firm age group-, size-, and sector-fixed effects. In column (1) of Table 11, a one-standard-deviation increase in the log of regional high school graduates increases the likelihood that a firm in the region engages in product innovations by 2.9 percentage points, which is a 6% increase in the average likelihood of product innovation (49%). The same increase in the log of high school graduates increases the likelihood of process and both types of innovations by 7.4 and 6.7 percentage points.[37]

To investigate which variation in the cross section of firms drives the results, I sort firms by the ratio of college-educated employees. Panel A of Figure 6 shows that across all three margins of innovation, the association decreases monotonically with the ratio of college-educated employees. I also run a double-interaction analysis, sorting firms by the ratio of college-educated employees and by the Pavitt sectorial taxonomy, which is based on the technological intensity of sectors (Figure A.7 of the Online Appendix). Firms that employ more basically-educated workers *and* operate in the least technologically intensive sectors drive the results.

Firms that innovate more invest more (Kogan et al. (2014)). For instance, they need to invest in property, plant, and equipment to buy the machines that allow producing their new products. In column (4) of Table 11, a one-standard-deviation increase in the log of high-school graduates is associated with a higher ratio between capital expenditures and previous end-of-year assets by 0.11, which is 4% of the average capital expenditures. I then split the sample of firms between those that innovate and those that do not innovate across the three margins (product, process, both types of innovations). Panel B of Figure 6 shows that firms that innovate across all margins drive the effect of basic education on firm-level investment. This result addresses the concern that basic education may capture regional determinants of investment unrelated to innovation.

In the sample, capital expenditures such as machines, property, and equipment are more likely to be financed with long-term debt (see Figure A.6 in the Online Appendix)). Hence, if higher basic education increases capital expenditures, it should also increase

[37]The results are similar if I estimate without restricting the variation within sectors, within firm age groups, and within firm size groups. The estimated coefficients corresponding to columns (1)-(3) of Table 11 are 2.6 p.p. (s.e. 1.4 p.p.), 7.4 p.p. (s.e. 1.4 p.p.), and 6.8 p.p. (s.e. 1.2 p.p.), respectively.

the ratio between long-term debt and total debt of firms. In column (5) of Table 11, a one-standard-deviation increase in the log of high school graduates in a region increases the ratio between long-term debt and total debt by about 16 percentage points. The association increases monotonically with the capital expenditures of firms, consistent with the notion that firms need to raise more long-term debt when they invest in capital expenditures to start new product lines (Panel C of Figure 6).

In Table 12, I run the IV analysis using the within-country distance from Mainz as an instrument for historical basic education to explain present-day firm-level outcomes. The effects of regional historical literacy on the present-day innovation, investment, and financing of firms in the IV analysis are similar to the reduced-form estimated effects, which I report in Table A.1 of the Online Appendix.

9 Alternative Channels and Explanations

In this section, I discuss a set of channels different from basic education, and alternative explanations of the results in the paper.

I first check if current or historical dimensions that correlate with historical literacy drive the regional associations of historical literacy with present-day regional innovation.[38] In Table 13, I consider (i) the ratio of high school-educated inhabitants in the region, (ii) the generalized trust index from the World Value survey at the regional level, (iii) the current dispersion of GDP per capita within regions, which captures income inequality in regions, (iv) the urbanization rate of the region around 1880 to proxy for economic performance at the time when historical literacy is measured, and (v) the index for the quality of past regional institutions proposed by Tabellini (2010). I add each dimension separately as a covariate in regressions of regional patents on regional literacy in 1880 and regional controls. Adding the percentage of inhabitants with a high school degree in column (2) of Table 13 decreases the estimated reduced-form effect of historical literacy on regional innovation, which stays marginally significant. Adding each of the other dimensions does not change the effect of historical literacy on innovation significantly. The results suggest that basic education may be a channel that transmits the effect of

[38]**??** of the Online Appendix tests for the correlation of a set of historical and current variables with historical literacy.

historical basic education on innovation.

Blue collars or machines? The automation of skilled and unskilled jobs has impacted investment, wages, and income distributions worldwide. Regions and firms with more blue-collar workers may be those with a higher scope for automation of jobs. As workers were substituted by robots, manufacturing firms may have had more investment opportunities. Thus, one may worry about a "reverse basic-education effect": the machines that substituted for the blue-collar workers are what improves the innovation of firms in traditional industries.[39] To assess this interpretation, I test for the effects of historical literacy on regional patents for each year during the period when Europe moved from no automation to its highest levels of automation, that is, from the late 1970s to the mid-1990s (Alesina and Zeira, 2006). Under the "reverse basic-education effect," the effect should increase in size over time: as machines are introduced at a faster pace, new innovation and investment opportunities arise for firms. Instead, under the basic-education interpretation, the effect should, if anything, decrease while machines substitute for blue-collar workers.[40] In Figure A.9 of the Online Appendix, the pattern of the association between basic education and regional innovation over time is not compatible with an increase in the size of the effect while the machines were introduced in Europe.

Blue collars or managers? Throughout the paper, basic education is interpreted as the education level of blue-collar workers. One may wonder if variation is present in the amount of basic education across managers, although the variation is plausibly larger among blue-collar workers than managers. In the EU-EFIGE/Bruegel database, I do not observe the education level of blue-collar workers and managers separately. I build on Bloom and Van Reenen (2007), who find the managerial practices of family firms that hire professional managers are slightly better than those of the average firm. Instead, the managerial practices of family firms that hire managers within the family are similar to those of the average firm, and they are worse if families use the primogeniture rule. If basic education improves managerial practices unrelated to blue-collar workers' characteristics, the effect should be larger for family firms that hire professional managers. But I find no significant difference in the size of the effects of basic education on innovation

[39]This interpretation is not easy to reconcile with the results on the intensive margin of basic education.

[40]In fact, the machines may also enable the blue collars that work with them to contribute to the innovation process.

if I estimate the specification of column (3) of Table 11 separately for family firms with professional managers (0.078, s.e. 0.039) or family managers (0.069, s.e. 0.015). The paper is agnostic on the mechanisms through which blue-collar workers improve the innovation of manufacturing firms. Better-educated blue-collar workers may propose product and process innovations to the management, or they may be better able to implement and experiment with innovative ideas coming from the management. Moreover, different skills of blue-collar workers may be important for innovation. Better scientific knowledge may help them develop innovative technical ideas, but better literary skills also allow them to express their ideas so that managers can understand them clearly.[41]

Financial constraints. One may be concerned that financial constraints alone explain all the results, because research has shown that financial constraints reduce the likelihood that firms innovate.[42] Akcomak and ter Weel (2009) find that early social capital increases current innovation across European regions, and they suggest the financing of innovation is easier in regions where social capital is stronger. The firm-level result that basic education increases the ratio of long-term debt over total debt signals that the channel this paper documents differs from the one studied by Akcomak and ter Weel (2009): if basic education acts only through financing, it should increase the overall debt of firms, whatever the debt's maturity. In Figure A.7 of the Online Appendix, I offer additional firm-level evidence that financial constraints alone cannot explain the results in this paper. I exploit the fact that the branches of business groups are less likely to be financially constrained than independent firms, because they may obtain funds from internal capital markets. Consistently, the effect of basic education on the likelihood of financial constraints is lower for branches than for independent firms. Instead, the effect of basic education on capital expenditures does not differ across the two groups. This effect should have been lower for branches if financial constraints alone drove it.

First Movers. Basic education may have caused some regions to engage in traditional manufacturing innovation first, and such regions may have perpetuated their primacy in

[41]Unfortunately, to guarantee the selection of schools participating in the program, run by regional authorities, is not biased toward the best regional schools, the region-level results of broad surveys based on standardized tests such as the PISA surveys are not diffused by most countries.

[42]But the evidence on the effects of financial constraints on innovation is mixed. For instance, Gorodnichenko and Schnitzer (2013) find constraints reduce innovation, whereas Almeida et al. (2013) find that they benefit firm-level innovation efficiency.

innovation.[43] This interpretation is at odds with a series of evidences in the economic history literature. Sandberg (1982) is among the first to show historical literacy across European countries is not correlated with historical GDP per capita, whereas it is positively correlated with GDP per capita today. Education seems to have been irrelevant to the take up of the Industrial Revolution in Europe (e.g., see Galor (2005) and Allen (2003)). Mokyr (2005a), Mokyr (2005b), and Mokyr and Voth (2009) propose that the upper tail of the knowledge distribution may have determined the early adoption of frontier technologies, and Squicciarini and Voigtlaender (2014) are the first to show that indeed the top of the distribution of education helped technology adoption and income growth in the First Industrial Revolution. Mass education, on the other hand, was less relevant.[44] The industrial organization literature also found results at odds with the first-mover interpretation. For instance, Raymond et al. (2010) find high-tech innovation is path-dependent, but innovation in traditional sectors is spuriously persistent: past innovation by itself does not cause future innovation. In this paper, the placebo distance results show the exogenous variation in the spatial distribution of past literacy I use is unrelated to the spatial diffusion of development or industrialization at the time the printing press was invented.

10 Conclusions

I study a neglected margin of innovation, the innovation of traditional manufacturing firms, and its effects on the investment and financing of firms. I find that a 10% increase in the number of high school graduates in European regions leads to 15% more patents filed in traditional manufacturing industries, and 4% higher investment in capital expenditures. I document these facts with a unique data set on the patented and unpatented innovation of European manufacturing firms. I construct Virtual Regions to show that unobservables that vary flexibly within and across regions do not drive the results. To address the issue of reverse causality for the effect of basic education on innovation, I use a shock to the acquisition of basic education by European households

[43]Literacy should not have also caused the same regions to engage in high-tech innovation first; otherwise, this interpretation is inconsistent with the evidence of no effect of basic education on high-tech innovation.

[44]The evidence on the Second Industrial Revolution is mixed. For instance, Becker et al. (2011) find literacy was important to the industrialization of Prussia.

before manufacturing started, and I verify the effects of the shock on basic education have persisted for decades.

The results suggest that learning-by-doing is not the only type of training that helps firms' innovation in traditional industries: formal basic education is also important. Hence, policies that address high school dropping may be relevant to firms that employ large shares of blue-collar workers. The results have additional policy implications. First, they may help explain why costly place-based transfers are often ineffective in increasing the productivity of firms in depressed areas, which are in most cases traditional manufacturing firms. Policies that address early school leaving may be more effective. An example is direct cash transfers to household heads and children conditional on school attendance (e.g., see Baird et al. (2011) and Bursztyn and Coffman (2012)). The results also have implications for the location decision of firms. U.S. and European firms that move their production overseas to reduce their wage and tax bills should account for the effect of higher basic education in their countries on investment opportunities.

Opening the black box of innovation in traditional industries suggests questions for future research in several fields, such as entrepreneurial finance, strategy, and development economics. For instance, how does the financing of high-tech innovation and of innovation in traditional manufacturing differ? How do managers' and workforce characteristics interact to produce innovation? And to what extent might manufacturing innovation in traditional sectors foster development in countries with no resources to engage in high-tech endeavors. The EU-EFIGE/Bruegel database used in this paper may also help address novel questions in corporate finance and trade, especially if referring to the activities of small and private firms across countries, which allows one to control for national culture and institutions.

This paper is one of the few in finance to study the long-term effect of historical shocks on financial outcomes. This approach may help financial historians enlarge the scope of their research, and may help finance scholars exploit the time dimension when looking for natural experiments to identify policy effects.

References

D. Acemoglu and D. Autor. What Does Human Capital Do? *Journal of Economic Literature*, 50(2):426–463, 2012.

D. Acemoglu, S. Johnson, and J. Robinson. The Colonial Origins of Comparative Development: an Empirical Investigation. *American Economic Review*, 91:1369–1401, 2001.

D. Acemoglu, F. Gallego, and J. Robinson. Institutions, Human Capital, and Development. *Annual Review of Economics*, 2014.

V. Acharya, R. Baghai, and K. Subramanian. Labor Laws and Innovation. *Journal of Law and Economics*, 56(4):997–1037, 2013.

P. Aghion, C. Meghir, and J. Vandenbussche. Growth, Distance to Frontier and the Composition of Human Capital. *Journal of Economic Growth*, 11:97–127, 2006.

I. Akcomak and B. ter Weel. Social capital, innovation and growth: Evidence from Europe. *European Econonomic Review*, 53:544–567, 2009.

A. Alesina and J. Zeira. Technology and Labor Regulations. *NBER Working Paper no. 12581*, 2006.

R. Allen. Progress and Poverty in Early Modern Europe. *Econonomic History Review*, 56(3):406–443, 2003.

H. Almeida, P. Hsu, and D. Li. Less is More: Financial Constraints and Innovation Efficiency. *Working Paper*, 2013.

C. Altomonte and T. Aquilante. The EU-Efige/Bruegel Dataset. *Bruegel Working Paper Series*, 2012.

P. Annoni and K. Kozovska. EU Regional Competitiveness Index 2010. *JRC Scientific and Technical Reports*, 2010.

S. Baird, C. McIntosh, and B. Ozler. Cash or Condition? Evidence from a Cash Transfer Experiment. *Quarterly Journal of Economics*, 126(4):1709–1753, 2011.

F. Barbier. L'Europe de Gutenberg: Le Livre et l'Invention de la Modernite' Occidentale. *Paris: Belin Editions*, 2006.

R. Barro. Economic Growth in a Cross-section of Countries. *Quarterly Journal of Economics*, 106:407–443, 1991.

Bureau of Economic Analysis BEA. Gross-Domestic-Product-by-Industry data, 1997-2013. 2014.

G. Becker. Investment in Human Capital: a Theoretical Analysis. *Journal of Political Economy*, 70(5):9–49, 1962.

S. Becker, E. Hornung, and L. Woessmann. Education and Catch-up in the Industrial Revolution. *American Economic Journal: Macroeconomics*, 3(3):92–126, 2011.

N. Bloom and J. Van Reenen. Measuring and Explaining Management Practices Across Firms. *QuarterlyJournal of Economics*, 122(4):1351–1408, 2007.

L. Bursztyn and L.C. Coffman. The Schooling Decision: Family Preferences, Intergenerational Conflict, and Moral Hazard in the Brazilian Favelas. *Journal of Political Economy*, 120(3):359–396, 2012.

C. Cameron and D. Miller. Robust Inference with Clustered Data. *Handbook of Empirical Economics and Finance*, pages 1–28, 2011.

G. Carlino and W. Kerr. Agglomeration and Innovation. *NBER Working Paper 20367*, 2014.

A. Chatterji, E. Glaeser, and W. Kerr. Clusters of Entrepreneurship and Innovation. *NBER Working Paper 19013*, 2013.

A. Ciccone and E. Papaioannou. Human Capital, the Structure of Production, and Growth. *Review of Economics and Statistics*, 91(1):66–82, 2009.

T. Conley. "GMM Estimation with Cross-Sectional Dependence". *Journal of Econometrics*, 105(1):59–83, 1999.

F. D'Acunto, M. Prokopczuk, and M. Weber. Deep-Rooted Antisemitism Shapes Households' Investments. *Working Paper*, 2014.

M. Dell. GIS Analysis for Applied Economists. *Manuscript*, 2009.

M. Dell. The Persistent Effect of Peru's Mining Mita. *Econometrica*, 78(8):1863–1903, 2010.

K. Desmet and E. Rossi-Hansberg. Innovation in Space. *American Economic Review P. and P.*, 102(3):447–452, 2012.

J.E. Dittmar. Information Technology and Economic Change: The Impact of the Printing Press. *Quarterly Journal of the Economics*, 126(3):1133–1172, 2011.

R. Easterlin. Why isn't the Whole World Developed? *Journal of Economic History*, 41(1):1–17, 1981.

EuropeanCommission. A Recovery on the Horizon? Annual Report on Eropean SMEs. 2013.

A. Galetovic, S. Haber, and R. Levine. Patent Holdup: do Patent Holders Holdup Innovation. *Working Paper*, 2014.

O. Galor. From Stagnation to Growth: Unified Growth Theory. *Handbook of Economic Growth*, 1:171–293, 2005.

N. Gennaioli, R. La Porta, F. Lopez-de Silanes, and A. Shleifer. Human Capital and Regional Development. *Quarterly Journal of Economics*, 128(1):105–164, 2013.

E. Glaeser, R. LaPorta, F. Lopez-de Silanes, and A. Shleifer. Do Institutions Cause Growth? *Jurnal of Economic Growth*, 9(3):271–303, 2004.

E. Glaeser, S. Pekkala Kerr, and W. Kerr. Entrepreneurship and Urban Growth: an Empirical Assessment with Historical Mines. *Review of Economics and Statistics, forthcoming*, 2014.

C. Goldin and L. Katz. The Race Between Education and Technology. *Harvard University Press*, 2008.

P. Gompers. Optimal Investmen, Monitoring, and the Staging of Venture Capital. *Journal of Finance*, 50:1461–1489, 1995.

P. Gompers and J. Lerner. The Venture Capital Revolution. *Journal of Economic Perspectives*, 15(2):145–168, 2001.

Y. Gorodnichenko and M. Schnitzer. Financial Constraints and Innovation: Why Poor Countries don't Catch Up. *Journal of the European Economic Association*, 11:1115–1152, 2013.

E.A. Hanushek and L. Woessmann. Do Better Schools Lead to More Growth? *Journal of Economic Growth*, 17:267–321, 2012.

S. Kaplan and P. Stromberg. Financial Contract theory meets the Real World: an Empirical Analysis of Venture Capital Contracts. *Review of Economic Studies*, 70: 281–315, 2003.

W. Kerr, J. Lerner, and A. Schoar. The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. *Review of Financial Studies*, 27(1):20–55, 2014.

L. Kogan, D. Papanikolau, A. Seru, and N. Stoffman. Technological Innovation, Resource Allocation and Growth. *Working Paper*, 2014.

S. Kortum and J. Lerner. Assessing the Contribution of Venture Capital to Innovation. *RAND Journal of Economics*, 31:674–692, 2000.

J. Lerner and U. Malmendier. With a Little Help from my Random Friends: Success and Failure in Post-Business School Entrepreneurship. *Review of Financial Studies*, 26(10): 2411–2452, 2013.

J. Lerner, M. Sorensen, and P. Stromberg. Private Equity and Long-Run Investment: The Case of Innovation. *Journal of Finance*, 66(2):445–477, 2011.

S. Levitt, J. List, and C. Syverson. Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant. *Journal of Political Economy*, 121(4): 643–681, 2013.

R. Lucas. On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22:3–42, 1988.

G. Manso. Motivating Innovation. *Journal of Finance*, 66:1823–1869, 2011.

S. Michalopoulos. The Origins of Ethnolinguistic Diversity. *American Economic Review*, 102(4):1508–1539, 2012.

S. Michalopoulos and E. Papaioannou. Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, 81(1):113–152, 2013.

J. Mincer. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4):281–302, 1958.

J. Mokyr. Long-Term Economic Growth and the History of Technology. *Handbook of Economic Growth*, 1:1113–1180, 2005a.

J. Mokyr. The Intellectual Origin of Modern Economic Growth. *Journal of Economic History*, 65(2):285–351, 2005b.

J. Mokyr and J. Voth. Understanding Growth in Early Modern Europe. *Cambridge Economic History of Europe*, Cambridge:Cambridge University Press, 2009.

E. Moretti. The New Geography of Jobs. *New York: Houghton Mifflin Harcourt*, 2012.

E. Moretti and D. Wilson. State Incentives for Innovation, Star Scientists and Jobs: Evidence from Biotech. *Journal of Urban Economics*, forthcoming.

P. Moser. How do Patent Laws Influence Innovation? Evidence from the Nineteenth-Century World Fairs. *American Economic Review*, 95(4):1214–1236, 2005.

R. Nelson and E. Phelps. Investment in Humans, Technological Diffusion and Economic Growth. *American Economic Review*, 56(1):69–75, 1966.

N Nunn. Historical Development. *Handbook of Economic Growth*, 2013.

OECD. Education at a Glance. *OECD Publishing*, 2013.

Organization for Economic Cooperation Development. OECD. Oslo Manual - Guidelines for Collecting and Interpreting Innovation Data. 2005.

T. Opler and S. Titman. Financial Distress and Corporate Performance. *Journal of Finance*, 49(3):1015–1040, 1994.

L. Putterman and D. Weil. Post-1500 Population Flows and the Long-Run Determinants of Economic Growth. *Quarterly Journal of Economics*, 125(4):1627–1682, 2010.

N. Ramankutty et al. The Global Distribution of Cultivable Lands. *Global Ecology and Biogeography*, 11:377–392, 2002.

W. Raymond, P. Mohnen, F. Palm, and S. Schim van der Loeff. Persistence of Innovation in Dutch Manufacturing: is it Spurious? *Review of Economics and Statistics*, 92(3): 495–504, 2010.

L. Sandberg. Ignorance, poverty and economic backwardness in the early stages of European industrialization: Variations on Alexandre Gerschenkron's grand theme. *Journal of European Economic History*, 11(3):675–697, 1982.

A. Saxenian. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, 1994.

T. Schoellman. Education Quality and Development Accounting. *Review of Economic Studies*, 79(1):388–417, 2012.

J. Schumpeter. The Theory of Economic Development. *Harvard University Press*, 1911.

E. Spolaore and R. Wacziarg. How Deep Are the Roots of Economic Development? *Journal of Economic Literature*, 51(2):325–369, 2013.

M. Squicciarini and N. Voigtlaender. Human Capital and Industrialization: Evidence from the Age of Enlightenment. *Working Paper*, 2014.

J. Stock and M. Yogo. Testing for Weak Instruments in Linear IV Regression. *Identification and Inference for Econometric Models. New York: Cambridge University Press*, 2005.

G. Tabellini. Culture and Institutions: Economic Development in the Regions of Europe. *Journal of the European Economic Association*, 8(4):677–716, 2010.

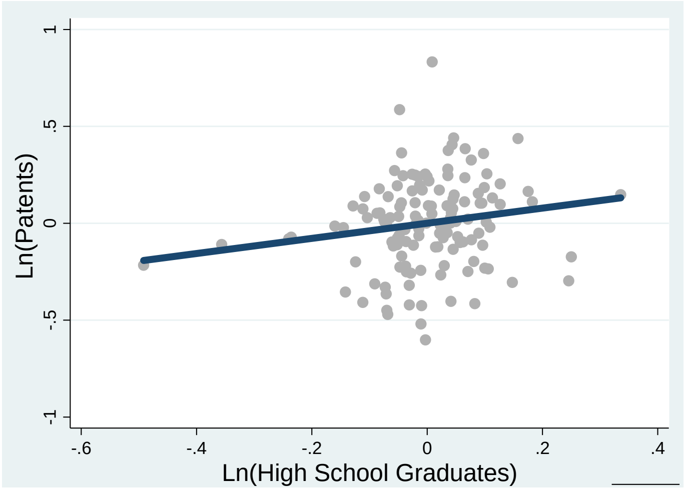
United States Patent Trademark Office USPTO. U.S. Patenting Trends by NAICS Industry Categories, 1963-2012. 2012.

J.A. Veblen. The Theory of Business Enterprise. *New York: Charles Scribner's Sons*, 1904.

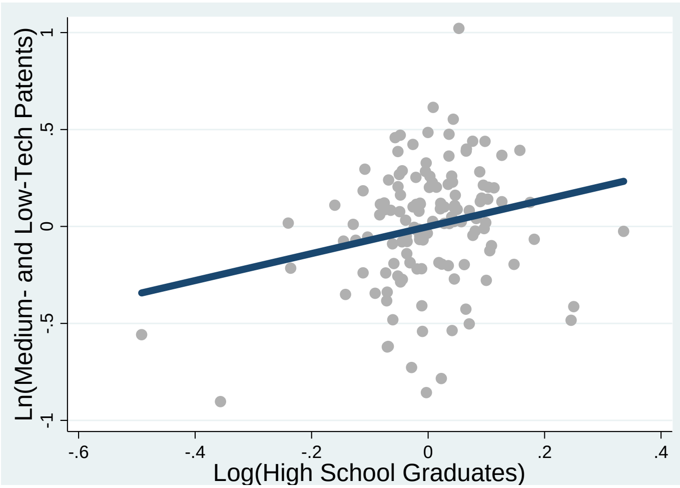
World Intellectual Property Organization WIPO. International Patent Classification - Colre Level, Volume 5. 2006.

Figure 1: Basic Education and Innovation in Regions

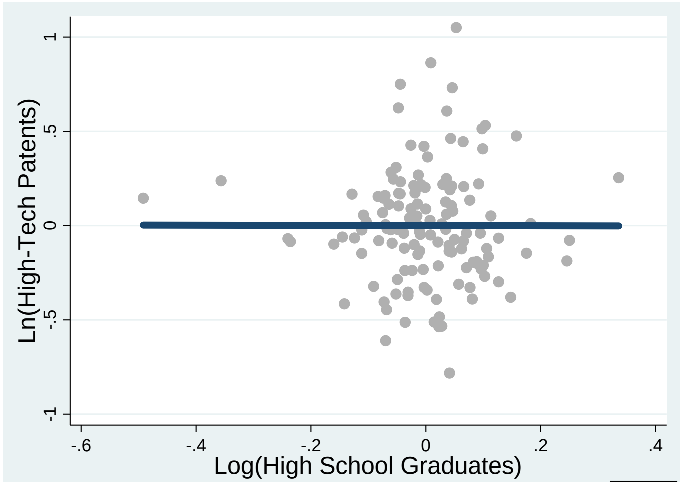
A. Basic Education and Patents in Regions



B. Basic Education and Low-Tech Patents



C. Basic Education and High-Tech Patents



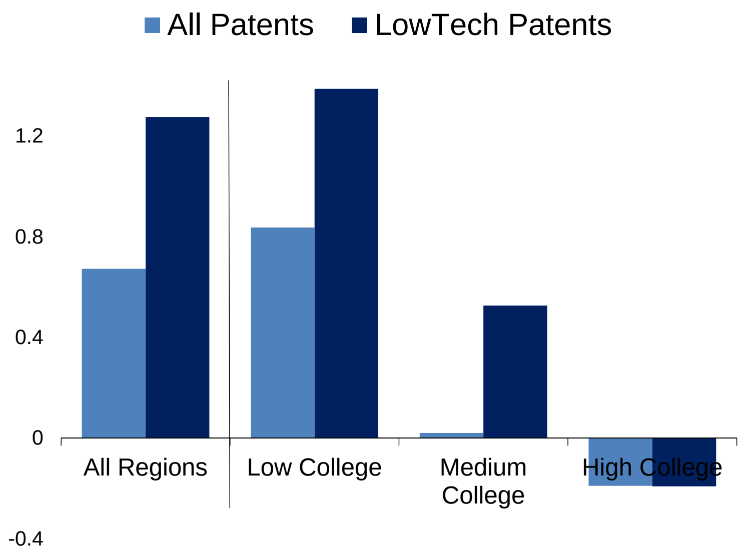
The graphs in Figure 1 plot the residuals from the following specification, where the dependent variables are regional patents in Panel A, regional low-tech patents in Panel B, and regional high-tech patents in Panel C:

$$Ln(Patents)_{r,c} = \alpha + \gamma Ln(College)_{r,c} + \theta Ln(NoDegree)_{r,c} + Ln(X)'_{r,c}\delta + \eta_c + \epsilon_{r,c},$$

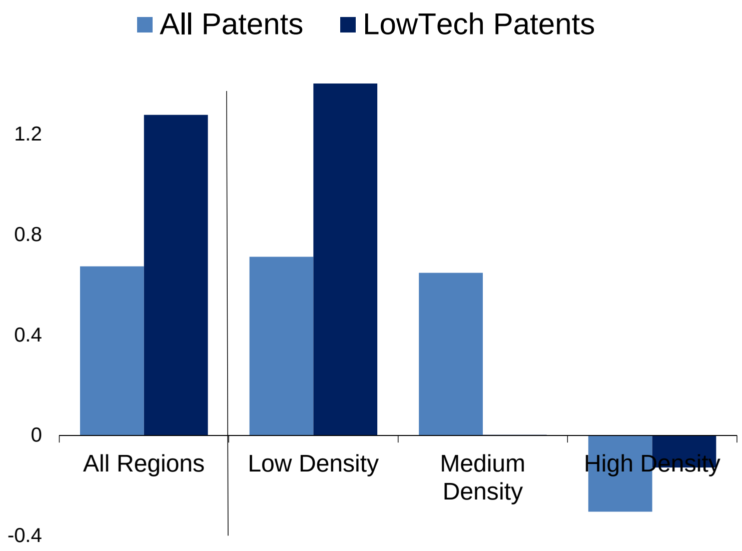
against the residuals from an analogous regression whose dependent variable is the log of high-school graduates ($Ln(HighSchool)_{r,c}$) in regions. X is a set of regional controls that include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002). The split of patents count by technological intensity follows the sectorial taxonomy by the International Patent Office, adopted by Eurostat. The following sectors are categorized as high-tech: Aviation, Computer, Communication Technology, Lasers, Micro-organism and Genetic Engineering, Semi-conductors.

Figure 2: INTERACTION EFFECTS - REGIONS

A. Effect of high school graduates on patents by share college educated



B. Effect of high school graduates on patents by population density



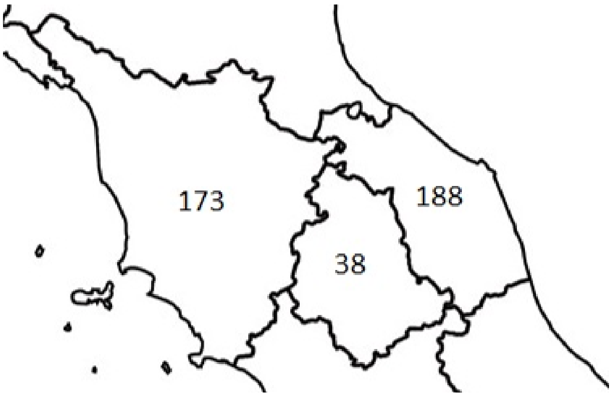
Panels A and B of Figure 2 plot the coefficients from the following OLS specification:

$$Ln(Patents)_{r,c} = \alpha + \beta Ln(HighSchool)_{r,c} + \gamma Ln(College)_{r,c} + Ln(X)'_{r,c}\delta + \eta_c + \epsilon_{r,c},$$

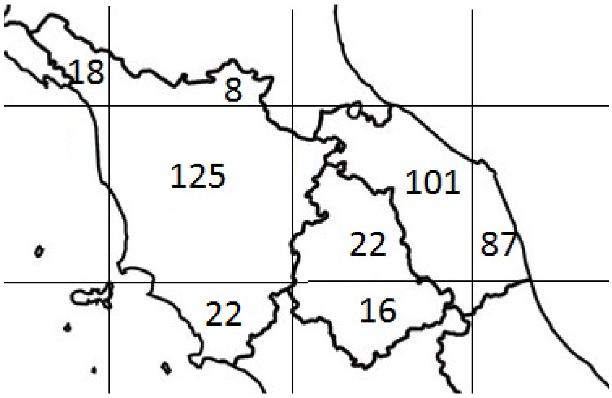
where $Ln(Patents)_{r,c}$ is the log of all patents filed in region r of country c in 2005 (light blue), or the log of mid/low-tech patents (dark blue). $Ln(HighSchool)_{r,c}$ is the log of inhabitants of region r of country c holding a high school degree as of 2005. X is a set of regional controls that include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002). η_c are a set of country fixed effects. In Panel A, regions are sorted in three equal-size groups based on the share of college-educated inhabitants in the region, and the specification is estimated within each group. In Panel B, regions are sorted in three equal-size groups based on the population density in the region, and the specification is estimated within each group.

Figure 3: Virtual Regions

A. Regional Patents



B. Split of County-level Patents within Regions



C. Sum of Patents within Virtual Regions

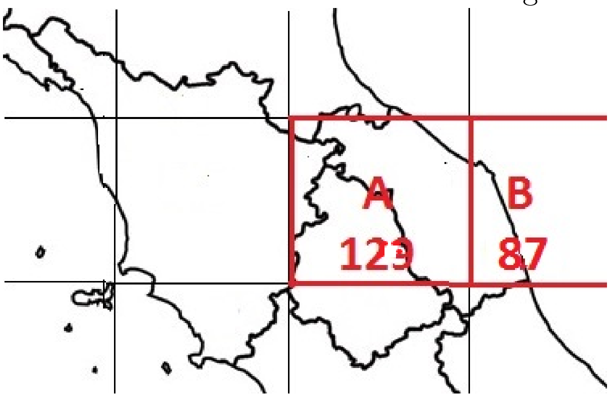


Figure 3 describes the construction of the Virtual-Region dataset for the variable Patents across three Italian central regions. A grid of size 100 Km by 100 Km is imposed to the map of Europe projected with an equal-area cylindrical projection. Each non-sea portion of a square is a Virtual region. Panel A reports the number of Patents at the regional level for the three regions in the example. Panel B shows how the county-level Patents of each region are assigned to the portions of the regions that enter different Virtual Regions: the county-level values are added across all the counties that enter the same Virtual Region. If a county is split between Virtual Regions, I assign the county value to the Virtual Region that covers its largest part. Panel C shows how the regional portions of Patents are aggregated at the level of Virtual Regions, by summing up the Patents for each regional partition in the Virtual Region.

Figure 4: Diffusion of the Printing Press after 1450

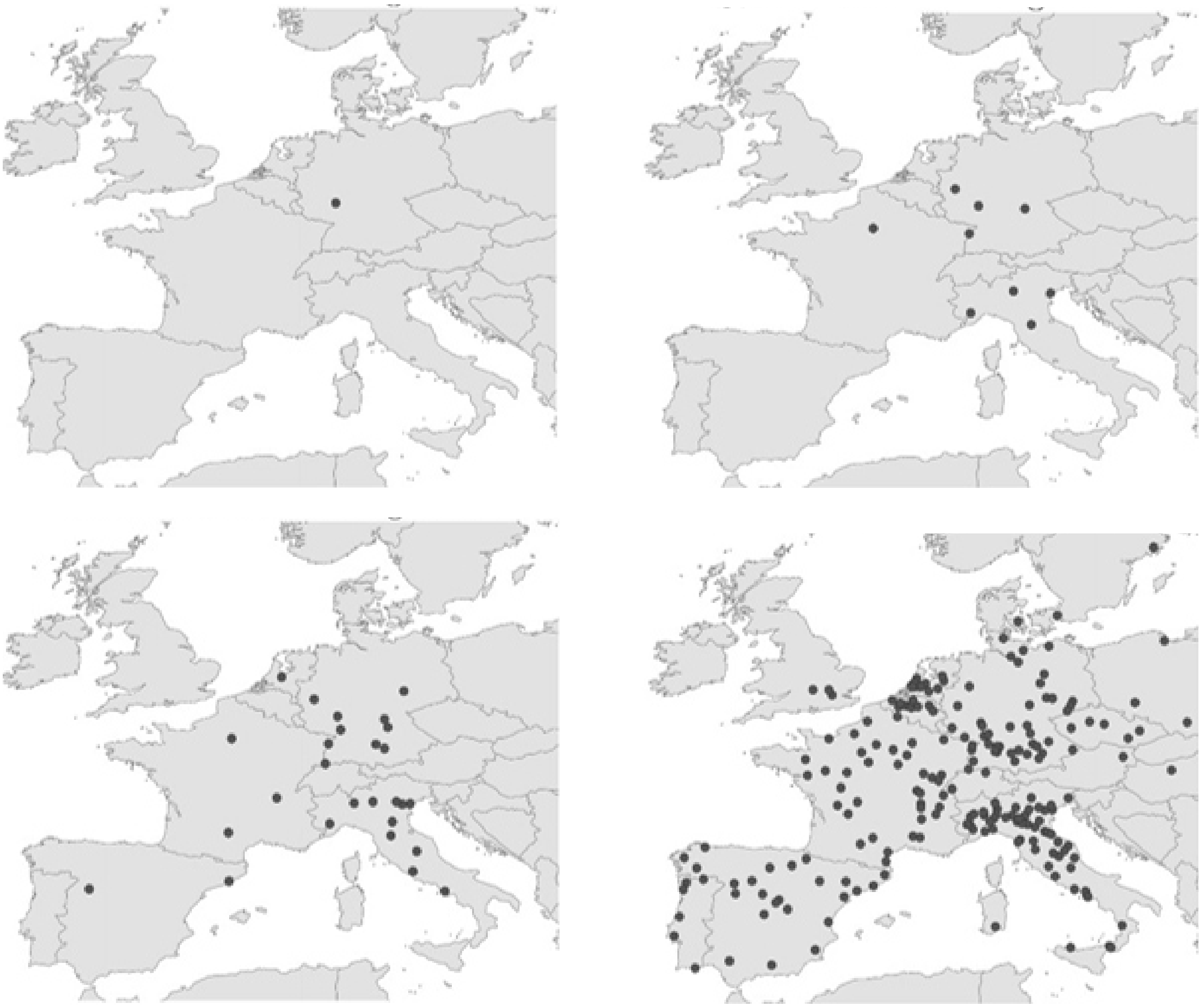
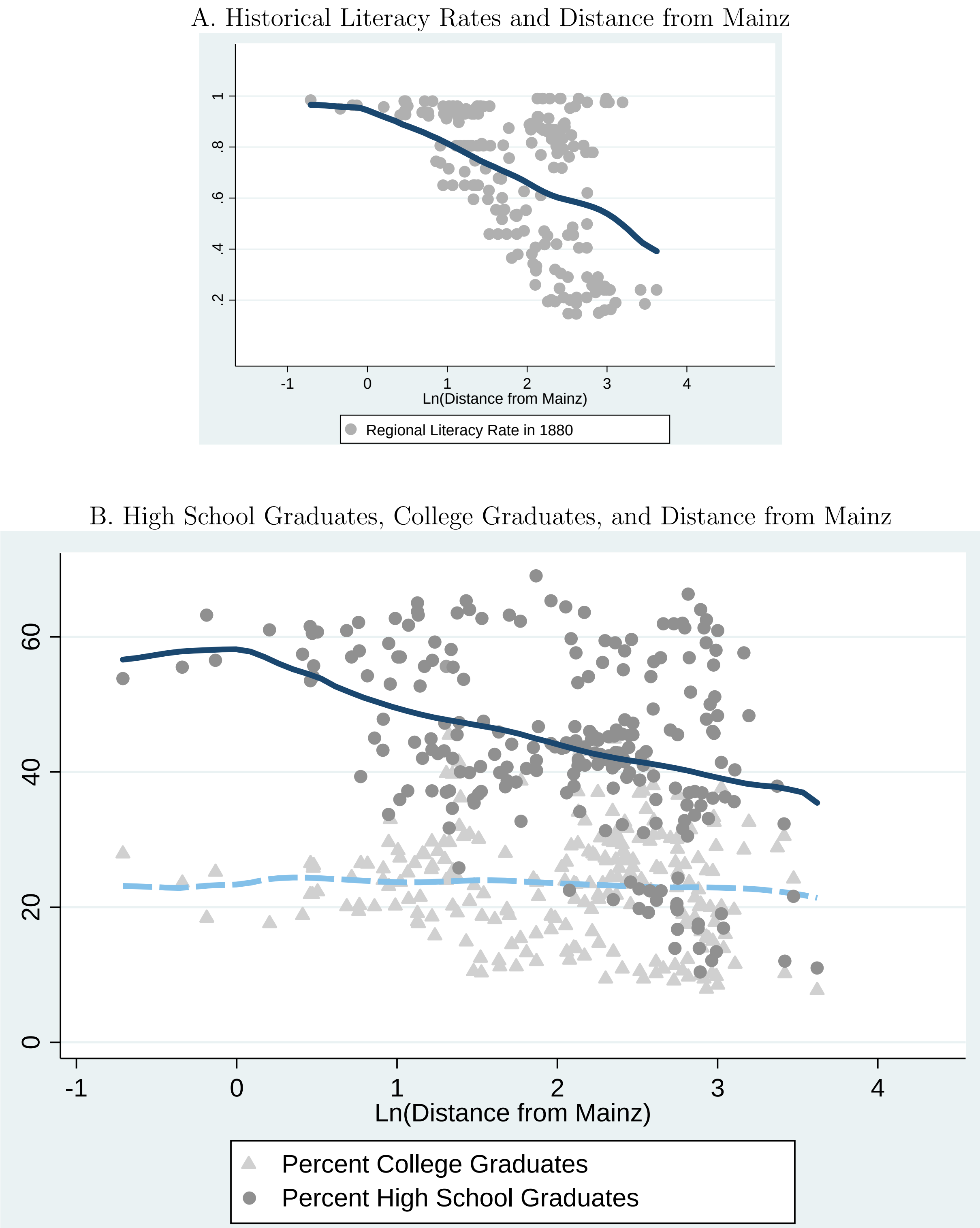


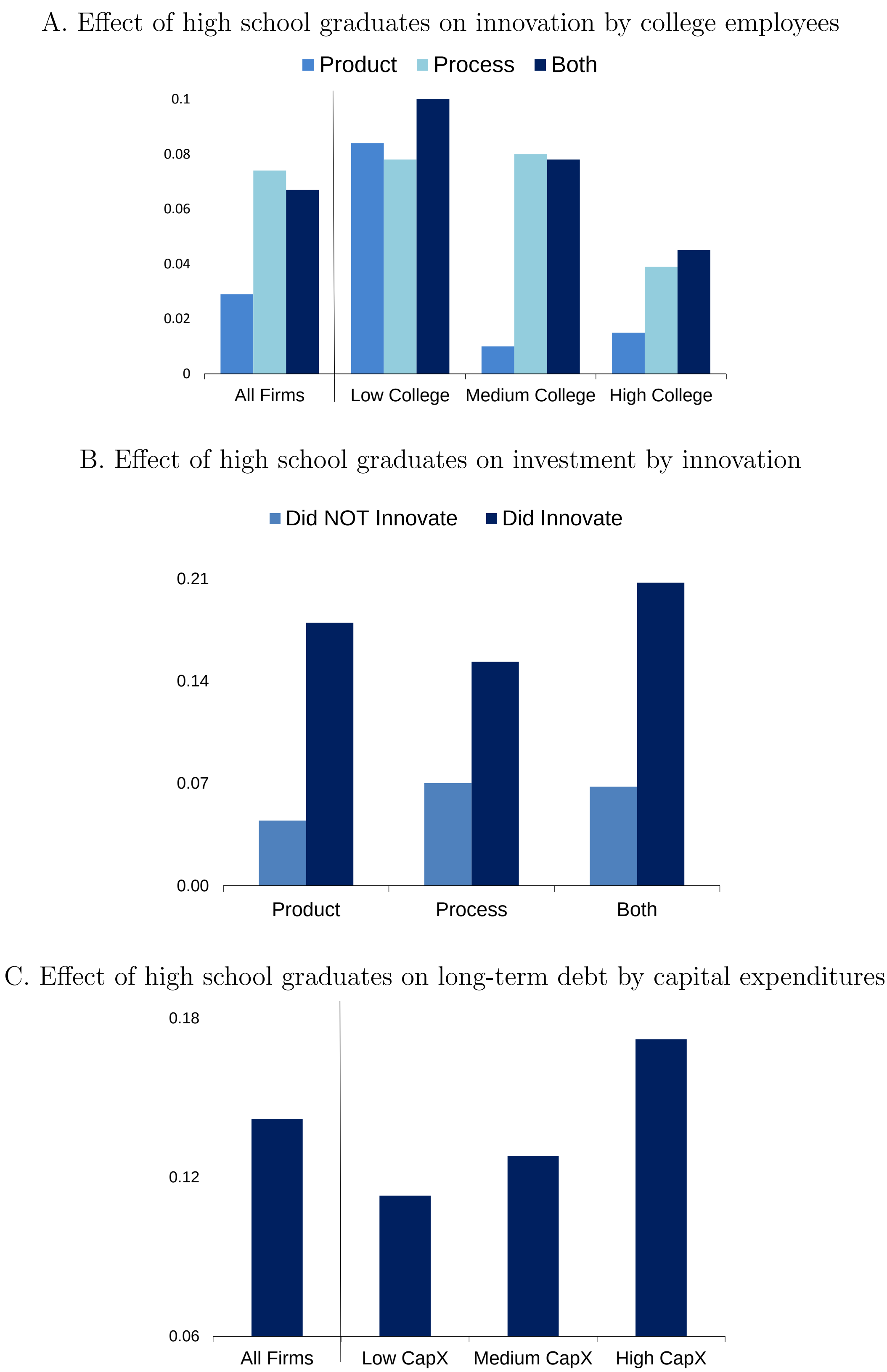
Figure 4 is a one-to-one replication of Figure III of Dittmar (2011). Each point on the maps represent a town where a printing press existed after its invention in 1450 in Mainz (current Germany), which is the only dot in the top-left graph of Figure 4. Each map describes the spatial diffusion of the printing press in the decades after it was invented.

Figure 5: Distance from Mainz, Literacy, and Current Education



Panel A of Figure 5 plots the unconditional correlation of the literacy rate in a European region as of 1880 against the log of the distance of the centroid of the region from the city of Mainz, in current Germany. Each point is a European region for which the historical literacy data is available. The dark blue line is a local linear polynomial fit for the relationship with a bandwidth of 1. Panel B of Figure 5 plots the unconditional correlation of the regional percentage of high school graduates (dark points), and of college graduates (light triangles) against the log of the distance of the centroid of the region from the city of Mainz, in current Germany. The solid dark blue line is a local linear polynomial fit with a bandwidth of 1 for the relationship of the percentage of high school graduates and the distance from Mainz. The dashed light blue line is a local linear polynomial fit with a bandwidth of 1 for the relationship of the percentage of college graduates and the distance from Mainz.

Figure 6: Interaction Effects - Firms



Panel A of Figure 6 plots the estimated coefficient $\hat{\beta}$ from the following probit specification run across three equal-size groups of firms in the EU/EFIGE-Bruegel data set sorted by the share of college-educated employees in the firm:

$$Pr(Innovation = 1)_{frc} = \Phi(\alpha + \beta Ln(HighSchool)_{rc} + X'_{rc}\gamma + F'_{frc}\delta + \eta_c + \eta_a + \eta_s + \eta_l), \tag{11}$$

where $Innovation$ is a dummy equal to 1 if the firm declares it engages in product, process, or both types of innovation. X_{rc} and F_{frc} are region- and firm-level covariates, and η_a, η_s, and η_l are firm age group-, size-, and sector-fixed effects. Panel B of Figure 6 plots the estimated coefficient on $Ln(HighSchool)$, the log of regional inhabitants with a high school degree as of 2005, in a OLS regression whose outcome is the capital expenditures of the firms in the EU/EFIGE-Bruegel data set normalized by previous end-of-year assets, and the RHS is the same as in Equation 11. The coefficient is estimated separately for firms that did not innovate in the two years before the survey was run (light blue), and firms that did innovate (dark blue). Panel C of Figure 6 plots the estimated coefficient on $Ln(HighSchool)$, the log of regional inhabitants with a high school degree as of 2005, in a OLS regression whose outcome is the share of long-term debt over total debt of the firms in the EU/EFIGE-Bruegel data set, and the RHS is the same as in Equation 11. The coefficient is estimated separately across three equal-size groups of firms sorted by the capital expenditures of the firm normalized by previous end-of-year assets.

Table 1: Summary statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	All Available information			Non-missing information		
Panel A. Regions	Obs	Mean	St. dev	Obs	Mean	St. dev
Current Characteristics						
Latitude	289	48.16	8.370	228	48.73	5.917
Population 05 (thousands)	289	1768	1465	228	1893	1565
GDP 05 (million €)	288	12824	40261	228	15591	44833
Patents 05	289	172.5	350.7	228	211.6	383.5
% High School 05	274	0.479	0.150	227	0.465	0.151
% College degree 05	274	0.225	0.081	227	0.230	0.078
Avg. Years School Basic 05	199	11.24	1.488	162	11.14	1.386
Avg. Years School College 05	225	16.78	4.121	184	16.81	4.123
Land Quality Index	280	0.592	0.255	225	0.584	0.238
Competitiveness Index	267	0.205	0.932	217	0.378	0.782
Trust WVS	234	0.323	0.151	185	0.325	0.137
Communist	289	0.170	0.376	228	0.145	0.353
Historical Characteristics						
Literacy 1880	228	0.698	0.263	228	0.698	0.263
Urbaniz. Rate 1860-1880	156	0.132	0.139	155	0.133	0.139
Historical Institutions index	156	0.458	2.177	155	0.468	2.180
Panel B. Firms	Obs	Mean	St. dev	Obs	Mean	St. dev
Financials/Ownership						
Total Assets (million €)	12538	14.90	134.1	12458	15.06	135.0
Sales (million €)	10620	22.41	205.3	10594	22.40	205.1
Cash flows (million €)	9449	1.076	17.10	9418	1.074	17.23
Property, Plant, Equipment (million €)	10612	1.919	25.30	10586	1.919	25.31
Tangibility	12434	0.255	0.201	12359	0.255	0.200
Employees	9342	116.6	4029	9313	116.6	4035
Family firm	14760	0.702	0.457	14185	0.709	0.454
Family CEO	14760	0.622	0.485	14185	0.628	0.483
Part of business group	14760	0.185	0.388	14185	0.185	0.388
Exports goods	14760	0.580	0.494	14185	0.579	0.494
Cut investments in 2010	12513	0.429	0.495	12056	0.428	0.494
Innovation						
Product Innovation	14760	0.491	0.499	14185	0.493	0.495
Process Innovation	14760	0.318	0.466	14185	0.322	0.467
Financial constraints						
Declares financial constraints	14760	0.174	0.379	14185	0.176	0.381
Rejected credit application	14760	0.040	0.196	14185	0.040	0.197
Asked for personal guarantees	14760	0.148	0.355	14185	0.148	0.356
Short term/Total debt	11287	0.406	0.394	11224	0.406	0.393

Table 1 reports summary statistics for the variables in the analysis. In both Panels, columns (1)-(3) provide statistics for all available observations, and columns (4)-(6) for regions and firms for which the historical literacy rates are observed. Panel A reports statistics for variables computed at the level of European regions (NUTS 2). Geographic and demographic variables are from the Eurostat Regional Database, and refer to 2005, the year employed in the cross-sectional analysis of Sections 4, 5, and 6. The regional count of patents filed in 2005 is from the Patstat-Kites database, based on the European Patent Office database. The average amount of years of schooling for those with basic education and college education are computed from the respondent to the World Value Survey (Wave 9), run from 1999 to 2004. This measure is not available for the regions of Austria, Belgium, Denmark, Greece, and Portugal. I measure the index of the quality of cultivable land at the regional level averaging the underlying 1 by 1 degree raster data of Ramankutty et al. (2002). I collect the first-wave results of the *EU Regional Competitiveness Index - RCI*, described in detail in Annoni and Kozovska (2010); I use the components on the quality of institutions and the quality of infrastructures. To obtain a regional measure of generalized trust, I average the individual level responses from the World Value Survey (Wave 9) at the regional level. Historical Literacy rates in regions as of 1880 are from Tabellini (2010), and for the primary sources (national censuses) for the regions of Greece and Spain. The literacy rate is the ratio of residents who could read and write in a European region around 1880. The historical urbanization rate of regions in the period 1860-1880, and the index of the quality of historical institutions are from Tabellini (2010). Panel B of Table 1 reports the statistics for the firms in the EU-EFIGE/Bruegel database, based on a survey run between 2008 and 2010 across seven European countries (Austria, Italy, Spain, UK, Germany, France, Hungary). Among the financials, Tangibility is measured as the ratio of tangible assets over total assets. For innovation, I create two dummies that equal 1 if a firm declares it engaged in any product or process innovations in the year prior to the interview. The definitions of product and process innovations are from the *Oslo Manual* (OECD, 2005): a product innovation is "the introduction of a good or service that is new or significantly improved with respect to its characteristics or intended uses." A process innovation is "the implementation of a new or significantly improved production or delivery method." The financial constraints variables are a set of dummies for whether the firms declare they faced financial constrained in the forms enlisted in the Table.

Table 2: Basic Education and Regional Innovation

Panel A. Extensive Margin of Basic Education	(1)	(2)	(3)	(4)	(5)	(6)
	All Patents		HighTech Patents		Mid/LowTech Patents	
Log Pop. High School	1.388***	1.008**	0.418	-0.331	2.377***	1.667***
	(0.508)	*(0.433)*	*(0.535)*	*(0.480)*	*(0.794)*	*(0.848)*
Log Pop. College	2.409***	1.382***	2.857***	1.925***	1.776***	0.639
	(0.392)	*(0.356)*	*(0.335)*	*(0.391)*	*(0.521)*	*(0.542)*
Regional controls	X	X	X	X	X	X
Country f.e.	X	X	X	X	X	X
Competitiveness Index, GDP p.c.		X		X		X
Observations	247	247	247	247	247	247
N. of clusters	88	88	88	88	88	88
Adjusted R^2	0.903	0.923	0.856	0.873	0.868	0.888

Panel B. Intensive Margin of Basic Education	(1)	(2)	(3)	(4)	(5)	(6)
	All Patents		HighTech Patents		Mid/LowTech Patents	
Log Avg.Years Basic Edu	1.733***	1.443**	1.274*	1.271	1.732**	1.219**
	(0.643)	*(0.577)*	*(0.645)*	*(0.819)*	*(0.799)*	*(0.590)*
Log Pop. Basic Edu	1.911	3.167	-3.596	-0.529	7.489***	8.710***
	(2.340)	*(2.438)*	*(2.206)*	*(2.783)*	*(3.044)*	*(2.997)*
Log Avg. Years Top Edu	-0.159	0.046	0.023	0.253	-0.032	0.145
	(0.370)	*(0.334)*	*(0.396)*	*(0.403)*	*(0.481)*	*(0.452)*
Log Pop. Top Edu	1.952***	2.304***	1.250	1.620**	3.024***	3.363***
	(0.654)	*(0.724)*	*(0.750)*	*(0.787)*	*(0.847)*	*(0.843)*
Regional controls	X	X	X	X	X	X
Country f.e.	X	X	X	X	X	X
Competitiveness Index, GDP p.c.		X		X		X
Observations	168	168	168	168	168	168
N. of clusters	62	62	62	62	62	62
Adjusted R^2	0.878	0.890	0.816	0.828	0.823	0.832

Panel A of Table 2 reports the estimated coefficients for the following OLS specification (extensive margin of basic education):

$$Ln(Patents)_{r,c} = \alpha + \beta Ln(HighSchool)_{r,c} + \gamma Ln(College)_{r,c} + Ln(X)'_{r,c}\delta + \eta_c + \epsilon_{r,c}, \tag{12}$$

where $Ln(Patents)$ are all regional patents in columns (1)-(2), high-tech regional patents in columns (3)-(4), and mid/low-tech regional patents in columns (5)-(6) filed in 2005. $Ln(HighSchool)_{k,r}$ is the log of inhabitants with a high school degree, and $Ln(College)_{k,r}$ is the log of inhabitants with a college degree that reside in region r of country c in 2005. Panel B of Table 2 reports the estimated coefficients for the following OLS specification (intensive margin of basic education):

$$\begin{aligned} Ln(Patents)_{r,c} = \alpha + \beta\theta Ln(s_{B,r,c}) + \beta Ln(BasicEducated)_{r,c} + \gamma\theta Ln(s_{C,r,c}) \\ + \gamma Ln(HigherEducated)_{r,c} + Ln(X)'_{r,c}\delta + \eta_c + \epsilon_{r,c}, \end{aligned} \tag{13}$$

where $s_{B,r,c}$ and $s_{C,r,c}$ are the average years of schooling of inhabitants with basic education and higher education, $Ln(BasicEducated)_{r,c}$ is the log of regional inhabitants with a high school degree or lower level of education, and $Ln(HigherEducated)_{r,c}$ is the log of regional inhabitants with more than high school education. The average years of schooling by education levels of regional inhabitants are computed from the respondents to the World Value Survey, which is not available for the full set of European regions. In each Panel, the sample size is restricted to the NUTS 2 regions for which all the regional observables are available. In both Panels, Regional controls include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002). Odd columns also include the *EU Regional Competitiveness Index - RCI* for the institutional quality and infrastructures components (Annoni and Kozovska, 2010), and the log of GDP per capita. In both Panels, standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 3: Robustness - Basic Education and Regional Innovation

.	(1)	(2)	(3)	(4)
	Extensive Margin of Basic Education		Intensive Margin of Basic Education	
	All Patents	Mid/Low Tech	All Patents	Mid/Low Tech
A. Excluding Top Patenting Regions	0.988	2.087	1.663	1.630
	*0.428***	*0.787****	*0.526****	*0.729***
Country f.e.	X	X	X	X
Other regional controls	X	X	X	X
Observations	233	233	155	155
N. of clusters	86	86	60	60
Adjusted R^2	0.913	0.868	0.872	0.802
B. Excluding Highest GDP Regions	1.712	3.201	1.315	1.313
	*0.499****	*0.736****	*0.591***	*0.775**
Country f.e.	X	X	X	X
Other regional controls	X	X	X	X
Observations	191	191	139	139
N. of clusters	80	80	58	58
Adjusted R^2	0.901	0.855	0.881	0.813
C. Excluding Lowest GDP Regions	1.153	1.876	2.776	2.379
	*0.532***	*0.951***	*1.052***	*1.159***
Country f.e.	X	X	X	X
Other regional controls	X	X		X
Observations	197	197	143	143
N. of clusters	84	84	59	59
Adjusted R^2	0.914	0.876	0.840	0.781
D. Negative Binomial Specifications	0.751	0.712	0.174	0.108
	*0.140****	*0.126****	*0.101**	*0.158*
Country f.e.	X	X	X	X
Other regional controls	X	X	X	X
Observations	247	247	169	169
N. of clusters	88	88	63	63
Unit covariate	1,000 ppl	1,000 ppl	1 year	1 year
E. Patents per capita and Ratios Graduates	0.318	0.380		
	*0.154***	*0.138****		
Country f.e.	X	X	–	–
Other regional controls	X	X		
Observations	247	247		
N. of clusters	88	88		
Adjusted R^2	0.588	0.514		

Panels A-C of Table 3 report the results for estimating log-log regressions of regional patents on the extensive margin (columns (1)-(2)) and the intensive margin (columns (3)-(4)) of basic education and regional controls across alternative subsamples. The outcome variables are all the patents filed in a region in 2005 (columns (1) and (3)), and the mid/low-tech patents filed in 2005 (columns (2) and (4)). In columns (1)-(2), the coefficients are those attached to the log of high school graduates in a region (extensive margin of basic education) in Equation 12 of Table 2. In columns (3)-(4), the coefficients are those attached to the log of the average number of years of schooling for individuals with high school degrees or lower levels of education (intensive margin of basic education) in Equation 13 of Table 2. The regional controls include the log of inhabitants without degrees, log of latitude, log of area, log of population density, the log of the index of the quality of cultivable land from Ramankutty et al. (2002), the *EU Regional Competitiveness Index - RCI* for the institutional quality and infrastructures components (Annoni and Kozovska, 2010), and the log of GDP per capita. Panels D-E estimate the relationship using specifications alternative to OLS. In Panel D, I estimate negative binomial regressions to account for the over-dispersed count nature of the patent data. In Panel E, I regress the patents per capita in a region on the ratio of inhabitants with a high school degree, the ratio of those with college degrees, and the other regional covariates. In all Panels, standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 4: REGIONAL UNOBSERVABLES: COUNTY-LEVEL ANALYSIS

	(1)	(2)	(3)	(4)
Log Pop. High School	1.522***	1.088**	1.124***	1.237***
	(0.421)	*(0.418)*	*(0.395)*	*(0.403)*
Log Pop. College	2.788***	1.956***	1.617***	1.603***
	(0.371)	*(0.380)*	*(0.401)*	*(0.406)*
County-level controls	X	X	X	X
Region f.e.		X	X	X
Excluding cities with >500,000 inhabitants			X	X
Excluding cities with >200,000 inhabitants				X
Observations	1,254	1,254	1,186	677
N. of clusters	92	92	87	65
Adjusted- R^2	0.48	0.86	0.93	0.96

Table 4 reports the estimated coefficient $\hat{\beta}_k$ in the following OLS specification:

$$Ln(Patents)_{k,r} = \alpha + \beta_k Ln(HighSchool)_{k,r} + \gamma_k Ln(College)_{k,r}$$
$$+Ln(X)'_{k,r}\delta + \eta_r + \epsilon_{k,r},$$

where $Ln(Patents)_{k,r}$ is the log of patents filed in county k of region r in 2005, $Ln(HighSchool)_{k,r}$ is the log of inhabitants with a high school degree in the county in 2005, $Ln(College)_{k,r}$ is the log of inhabitants with a college degree in the count in 2005, and η_r are regional fixed effects. County-level controls include the log of inhabitants without degrees, log of latitude, log of area, log of population density, the log of the index of the quality of cultivable land from Ramankutty et al. (2002), and the log of the GDP per capita, all measured at the county level. The *EU Regional Competitiveness Index - RCI* for the institutional quality and infrastructures components is not available at the county level. Standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 5: Unobservables within and across regions: Virtual Regions

	(1)	(2)	(3)	(4)
Log Pop. High School	1.157***	1.087**	1.213***	1.185***
	(0.348)	*(0.412)*	*(0.391)*	*(0.383)*
Log. Pop. College		2.063***		1.976***
		(0.438)		*(0.364)*
Country f.e.	X	X		
Region f.e.			X	X
Virtual-Region controls		X		X
Observations	560	560	560	560
N. of clusters	82	82	82	82
Adjusted- R^2	0.19	0.81	0.94	0.95

Table 5 reports the estimated $\hat{\beta_v}$ from the following specification:

$$Ln(Patents)_{v,r*} = \alpha + \beta_v Ln(HighSchool)_{v,r*} + \gamma_v Ln(College)_{v,r*}$$
$$+Ln(X')_{v,r*}\delta_v + \eta_{r*} + \epsilon_{v,r*},$$

where $Ln(Patents)_{v,r*}$ are the patents in Virtual Region v, indexed to the region r* that cover its largest part; η_{r*} is a set of regional fixed effects for the underlying $r*$ regions. Virtual Regions are obtained by imposing an arbitrary grid of 100 km by 100 km on the map of Europe. Virtual-Region variables are obtained by aggregating the values of the corresponding variables at the county level, for the counties that enter the Virtual Region. If a county is split across two or more Virtual Regions, I assign the county-level values to the Virtual Region that covers the largest part of the county. Virtual-Region controls include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002). The *EU Regional Competitiveness Index - RCI* for the institutional quality and infrastructures components is not available at the county level, hence it does not enter the Virtual-Region analysis. The number of regional fixed effects is 234, and the number of groups of regions is 82, because not all the regions of Europe are assigned to one or more Virtual Regions. Standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 6: Virtual Regions: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	No Top Patents		No smallest Regions		No High Population Density	
Log Pop. High School	1.162***	1.204***	1.209***	1.177***	1.344***	1.469***
	(0.388)	*(0.369)*	*(0.405)*	*(0.391)*	*(0.547)*	*(0.575)*
No Top 5%	X		X		X	
No Top 10%		X		X		X
Virtual-Region controls	X	X	X	X	X	X
Region f.e.	X	X	X	X	X	X
Observations	532	501	532	501	532	501
N. of clusters	79	76	80	78	79	75
Adjusted- R^2	0.90	0.91	0.93	0.94	0.94	0.94

Table 6 reports the estimated $\hat{\beta_v}$ from the following specification run on a set of subsamples of the full sample of Virtual Regions:

$$Ln(Patents)_{v,r*} = \alpha + \beta_v Ln(HighSchool)_{v,r*} + \gamma_v Ln(College)_{v,r*} \\ + Ln(X')_{v,r*}\delta_v + \eta_{r*} + \epsilon_{v,r*},$$

where $Ln(Patents)_{v,r*}$ are the patents in Virtual Region v, indexed to the region r* that cover its largest part; η_{r*} is a set of regional fixed effects for the underlying $r*$ regions. Virtual Regions are obtained by imposing an arbitrary grid of side 100 km on the map of Europe. All Virtual-Region variables are obtained by aggregating the values of the corresponding variables at the county level, for the counties that enter the Virtual Region. If a county is split across two or more Virtual Regions, I assign the county-level values to the Virtual Region that covers the largest part of the county. Virtual-Region controls include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002). Standard errors are clustered at the level of groups of regions (NUTS 1). The number of regional fixed effects is 234, and the number of groups of regions is 82, because not all the regions of Europe are assigned to one or more Virtual Regions. Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 7: Virtual Regions: Alternative Shapes and Positions of the Grid

Configurations: 4 grid shapes, 27 positions	
Number of configurations	108
Significant coefficients at the 5% level of significance	106 (98.1%)
Median coefficients	1.170
Mean coefficients	1.197
Standard deviation coefficients	0.249
Ratio Mean/Standard deviation	4.79
Mean standard errors	0.351
Standard deviation standard errors	0.042

Simulations: Summary coefficients by shape	
Squares	Triangles
Median: 1.173	Median: 1.175
Mean: 1.171	Mean: 1.167
Diagonal Squares	Parallelograms
Median: 1.162	Median: 1.251
Mean: 1.214	Mean: 1.236

Table 7 reports the summary of results for estimating the coefficient β_v in the following Virtual-region specification across 108 alternative configurations of the Virtual Regions:

$$Ln(Patents)_{v,r*} = \alpha + \beta_v Ln(HighSchool)_{v,r*} + \gamma_v Ln(College)_{v,r*} \\ + Ln(X')_{v,r*}\delta_v + \eta_{r*} + \epsilon_{v,r*},$$

The configurations use four shapes of the grid, each translated in 27 alternative positions. The shapes include: (i) 100 km by 100 km squares, like those in the analysis so far; (ii) isosceles triangles of side 100 km, obtained by dividing each square in half along its northwest to southeast diagonal; (iii) parallelograms of side 200 km and height 100 km, obtained by merging two triangles across different squares; and (iv) squares whose sides are the diagonals of the original squares. By construction, the four shapes produce Virtual Regions that generally differ from each other in their county composition. As for the positions of the grid, I translate each grid in 27 alternative ways to explore alternative positions with non-overlapping grids: (i) 9 times east, in increments of 10 km; (ii) 9 times north, in increments of 10 km; and (iii) 9 times northeast, in increments of 10 km * $\sqrt{2}$. Translated grids for each shape modify the composition of Virtual Regions as long as counties mainly covered by one square move to another square after the translation. Moreover, the grids produce alternative rules of regional indexing; hence, different groups of Virtual Regions enter the same fixed effect.

Table 8: PERSISTENCE OF BASIC EDUCATION: HISTORICAL LITERACY AND CURRENT EDUCATION

Panel A. Regions	(1)	(2)	(3)	(4)	(5)	(6)
Inhabitants by degree	Log High School		Log College		Log Immigrants	
Literacy 1880	0.049**	0.055**	0.005	-0.007	0.002	-0.007
	(0.023)	*(0.021)*	*(0.065)*	*(0.063)*	*(0.523)*	*(0.503)*
Regional controls	X	X	X	X	X	X
Country f.e.	X	X	X	X	X	X
Exclude no regional variation Literacy 1880		X		X		X
Observations	227	204	227	204	217	194
N. of clusters	78	70	78	70	78	70
(Adj., Pseudo) R^2	0.972	0.988	0.973	0.975	0.199	0.193

Panel B. Individuals	(1)	(2)	(3)	(4)	(5)	(6)
Years of schooling	All inhabitants		High School or lower		Some college	
Literacy 1880	0.921***	0.138	0.665***	0.196	0.045	0.077
	(0.140)	*(0.135)*	*(0.174)*	*(0.142)*	*(0.349)*	*(0.374)*
Individual controls	X	X	X	X	X	X
Regional controls	X	X	X	X	X	X
Country f.e.	X	X	X	X	X	X
Town size f.e.	X	X	X	X	X	X
Degree f.e.		X		X		X
Observations	5383	5383	4327	4327	1037	1037
N. of clusters	31	31	31	31	30	30
(Adj., Pseudo) R^2	0.204	0.580	0.228	0.374	0.083	0.083

Panel A of Table 8 reports results for estimating the following OLS specification at the level of regions (NUTS 2):

$$Ln(HighSchool)_{rc} = \alpha + \beta Literacy1880_{rc} + X'_{rc}\gamma + \eta_c + \epsilon_{rc},$$

where $Ln(HighSchool)_{rc}$ is the number of inhabitants with a high school but not higher degree in region r in country c; $Literacy1880_{rc}$ is the literacy rate in the region in 1880, whose sources are Tabellini (2010) and the national censuses of Greece and Spain for the regions not covered by Tabellini (2010); X_{rc} is a vector of regional controls that include the latitude, area, population density, GDP per capita, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002); η_c is a set of fixed effects at the country level.
Panel B of Table 8 reports results for estimating the following OLS specification at the level of household heads which take part for the first time in the World Value Survey from 1981 to 2010:

$$YearsSchooling_{irc} = \alpha + \beta Literacy1880_{rc} + X'_{rc}\gamma + D'_{irc}\delta + \eta_c + \epsilon_{irc},$$

where $YearsSchooling_{irc}$ is the number of years of schooling for household head i in region r and country c measured from the World Value Survey. X_{rc} is a vector of regional controls that include the latitude, area, population density, GDP per capita, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002); D_{irc} is a set of individual-level characteristics that include the gender, age (second polynomial), income (second polynomial); Town size f.e. are a set of 5 fixed effects for the size of the town-city where the individual resides. In both Panels, standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 9: Distance from Mainz: Exclusion Restriction

A. Distance from Mainz and Regional Observables

	(1)	(2)		(3)	(4)
Region-level observables	Plain Distance	Residual Distance	Firm-level observables	Plain Distance	Residual Distance
Ln(Pop. Density)	-0.090	0.288	Family Firm	-0.031	-0.047
	0.098	*0.215*		*0.031*	*0.025**
Ln(Self Employed)	0.047	0.068	Part of business group	0.020	0.022
	0.070	*0.109*		*0.026*	*0.019*
Ln(College Pop.)	-0.003	-0.090	Did Export	0.001	-0.003
	0.107	*0.124*		*0.028*	*0.024*
Regional Area	0.174	-0.756	Leverage	-0.148	-0.094
	0.161	*0.505*		*0.049***	*0.065*
Past Institution Quality Index	-0.150	-0.083	Tangibility	0.023	0.011
	0.192	*0.138*		*0.012*	*0.006*
Current Institution Quality Index	-0.147	0.011	Ln(Sales)	-0.080	0.005
	0.115	*0.113*		*0.048*	*0.023*
Country f.e.	X	X	Country f.e.	X	X
			Sector f.e.	X	X
			Age f.e.	X	X
			Size f.e.	X	X

B. Distance from Mainz and Historical Literacy in Reduced Form

Dep. Variable: Ln(Patents)			
Ln(Literacy 1880)	0.232		0.302
	*0.073****		*0.064****
Ln(Distance Mainz)		-0.170	-0.122
		*0.080***	*0.074*
Regional controls	X	X	X
Country f.e.	X	X	X
Observations	227	227	227
N. of clusters	78	78	78
Adjusted R^2	0.888	0.888	0.898

C. Placebo First Stages with other relevant Distances

	K-P F stat	A-P chi-sq	Correlation Dist. Mainz
Ln(Distance from Prague)	2.45	2.77	0.567
Ln(Distance from Amsterdam)	4.36	4.92	0.760
Ln(Distance from Madrid)	0.06	0.07	0.488
Ln(Distance from London)	3.90	4.40	0.390
Ln(Distance from Florence)	7.84	8.85	0.606
Ln(Distance from Aix-la-Chapelle)	7.71	8.70	0.881
Ln(Distance from Mainz)	10.51	11.86	1

Table 10: 2SLS: Instrumented Historical Literacy and Regional Innovation

Panel A. Second Stage	(1) All Regions	(2) No Communist	(3) No South	(4) No Germany	(5) No all three
Ln(Literacy1880)	0.574***	0.524***	0.616**	0.618***	0.640*
	(0.126)	*(0.123)*	*(0.273)*	*(0.131)*	*(0.320)*
Regional controls	X	X	X	X	X
Country f.e.	X	X	X	X	X
Observations	227	194	173	188	111
N. of clusters	78	66	59	63	37
Adj- R^2	0.877	0.869	0.868	0.847	0.714
Panel B. First Stage					
CD F-Statistic	48.66	42.26	16.38	79.13	35.91
KP F-Statistic	10.51	9.43	7.10	21.96	17.62
AP Chi-sq.	11.86	10.68	8.12	25.13	20.97

Panel A of Table 10 reports the estimated second-stage coefficients from two-stage least squares regressions whose outcome variable is the log of patents filed in a region in 2005, and the log of the literacy rate in the region in 1880 is instrumented with the log of the minimal Euclidean distance of the centroid of a region from the city of Mainz, in Germany. Regional controls include the log of the latitude, area, population density, GDP per capita, and of the index of the quality of cultivable land from Ramankutty et al. (2002). Each column refers to an alternative subsample or regions. Panel B of Table 10 reports the first-stage statistics for the two-stage least square analyses: (i) the Cragg-Donald F-statistic, which is based on i.i.d. standard errors, but is used to compute the critical values reported in Table 5.2. of Stock and Yogo (2005); (ii) the Kleibergen-Paap F-statistic, which is computed for correcting the standard errors for correlation of unknown form at the level of groups of regions; and (iii) the Angrist-Pischke chi-square statistic, which can be used for a rank test of the matrix of the reduced-form equation coefficients and the excluded instruments. Standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 11: Basic Education and Firm-Level Innovation, Investment, and Capital Structure

	(1)	(2)	(3)	(4)	(5)
		Innovation		Investment	Capital Structure
	Product	Process	Both	CapX	LT Debt/ Total Debt
Log Pop. High School	0.029**	0.074***	0.067***	0.135***	0.158***
	(0.013)	*(0.015)*	*(0.013)*	*(0.039)*	*(0.022)*
Log Pop. College	0.016*	0.012	0.011	-0.032	0.019
	(0.009)	*(0.012)*	*(0.009)*	*(0.028)*	*(0.015)*
Tangibility	0.001	0.031***	0.018***	0.048***	0.087***
	(0.006)	*(0.006)*	*(0.005)*	*(0.013)*	*(0.006)*
Leverage	0.015**	0.023**	0.022**	0.014*	0.001
	(0.007)	*(0.010)*	*(0.010)*	*(0.009)*	*(0.002)*
Log Sales	0.016**	0.038***	0.027***	0.232***	-0.022***
	(0.008)	*(0.007)*	*(0.007)*	*(0.022)*	*(0.007)*
Family firm	0.056**	0.038	0.042*	-0.008	-0.017
	(0.022)	*(0.024)*	*(0.021)*	*(0.049)*	*(0.017)*
Family CEO	-0.009	-0.006	-0.005	0.055	0.029*
	(0.020)	*(0.020)*	*(0.018)*	*(0.049)*	*(0.018)*
Part business group	-0.013	-0.011	-0.007	-0.122***	-0.062***
	(0.016)	*(0.015)*	*(0.015)*	*(0.032)*	*(0.016)*
Exports products	0.179***	0.050***	0.065***	-0.108***	-0.009
	(0.012)	*(0.010)*	*(0.010)*	*(0.026)*	*(0.009)*
Sector f.e.	X	X	X	X	X
Age f.e.	X	X	X	X	X
Size f.e.	X	X	X	X	X
Regional controls	X	X	X	X	X
Country f.e.	X	X	X	X	X
Observations	10140	10140	10140	10132	8083
N. of clusters	141	141	141	141	140
(Adj/Pseudo) R^2	0.062	0.020	0.028	0.101	0.174

Columns (1)-(3) of Table 11 report the results for estimating the following probit specification:

$$Pr(Innovation = 1)_{frc} = \Phi(\alpha + \beta Ln(HighSchool)_{rc} + X'_{rc}\gamma + F'_{frc}\delta + \eta_c + \eta_a + \eta_s + \eta_l) \tag{14}$$

where $Innovation$ is a dummy that equals 1 if the firms engaged in a product, process, or both types of innovations in the two years prior to participating in the Bruegel/EFIGE-Unicredit survey, which was run from 2008 to 2010; X is a set of Regional controls which include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002); F is a set of firm-level controls which include Tangibility (the ratio of tangible assets to total assets), Leverage (the ratio of total debt to shareholders' equity), the log of sales, and a set of dummy variables that equal one if the firm is owned by a family or entrepreneur, if the firm's CEO is a member of the owning family, if the firm is part of a business group, and if the firm engages in exporting of its products outside the country where it operated. Columns (4)-(5) of Table 11 report the results from estimating OLS regressions whose outcomes are the capital expenditures of the firm (property, plant, and equipment, normalized by previous end-of-year assets), and the ratio between long-term debt and total debt of the firm. The set of regional and firm-level controls are the same as those in Equation 14. Standard errors are clustered at the level of regions (NUTS 2). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 12: 2SLS: Instrumented Historical Literacy and Firm-level Outcomes

Panel A. Second Stage	(1)	(2)	(3)	(4)	(5)
		Innovation		Investment	Capital Structure
	Product	Process	Both	CapX	LT Debt/ Total Debt
Ln(Literacy 1880)	0.014*	0.020**	0.027***	0.113***	0.085***
	(0.007)	*(0.010)*	*(0.008)*	*(0.026)*	*(0.012)*
Regional controls	X	X	X	X	X
Firm-level controls	X	X	X	X	X
Country f.e.	X	X	X	X	X
Sector f.e.	X	X	X	X	X
Size f.e.	X	X	X	X	X
Age group f.e.	X	X	X	X	X
Observations	10,137	10,137	10,137	10,129	8080
N. of clusters	140	140	140	140	139
Adjusted R^2	0.088	0.029	0.029	0.116	0.241
Panel B. First Stage					
KP F-Statistic	68.02	68.02	68.02	68.03	78.31
AP Chi-sq.	68.72	68.72	68.72	68.73	79.18

Panel A of Table 12 reports the estimated second-stage coefficients from two-stage least squares regressions where the log of the literacy rate in a region is instrumented with the log of the minimal Euclidean distance of the centroid of a region from the city of Mainz, in Germany. In columns (1)-(3) *Innovation* is a dummy that equals 1 if the firms engaged in a product, process, or both types of innovations in the two years prior to participating in the Bruegel/EFIGE-Unicredit survey, which was run from 2008 to 2010; X is a set of Regional controls which include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002); F is a set of firm-level controls which include Tangibility (the ratio of tangible assets to total assets), Leverage (the ratio of total debt to shareholders' equity), the log of sales, and a set of dummy variables that equal one if the firm is owned by a family or entrepreneur, if the firm's CEO is a member of the owning family, if the firm is part of a business group, and if the firm engages in exporting of its products outside the country where it operated. In columns (4)-(5) of Table 12, the outcomes are the capital expenditures of the firm (property, plant, and equipment, normalized by previous end-of-year assets), and the ratio between long-term debt and total debt of the firm. The set of regional and firm-level controls are the same as those described above.
Panel B of Table 12 reports the first-stage statistics for the two-stage least square analyses: (i) the Kleibergen-Paap F-statistic, which is computed for correcting the standard errors for correlation of unknown form at the level of groups of regions; and (ii) the Angrist-Pischke chi-square statistic, which can be used for a rank test of the matrix of the reduced-form equation coefficients and the excluded instruments. Standard errors are clustered at the level of regions (NUTS 2). Statistical significance is shown as follows: ***1%, **5%, *10%.

Table 13: ALTERNATIVE CHANNELS

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Ln(Patents)					
Ln(Literacy 1880)	0.349	0.162	0.392	0.380	0.346	0.284
	*0.094****	*0.090**	*0.090****	*0.097****	*0.150***	*0.125**
Ln(% High School)		0.272				
		*0.070****				
Ln(% Generalized Trust)			0.042			
			0.046			
Ln(Dispersion GDP within region)				-0.064		
				0.044		
Ln(Urbanization rate 1860-1880)					-0.126	
					*0.024****	
Ln(Quality Historical Institutions)						-0.263
						*0.084***
Regional controls	X	X	X	X	X	X
Observations	224	224	184	216	152	161
R^2	0.846	0.864	0.850	0.850	0.730	0.873

Each column of Table 13 refers to a different OLS specification of the following form:

$$Ln(Patents)_{r,c} = \alpha + \beta Literacy1880_{r,c} + \gamma RegCovar_{r,c} + X'_{r,c}\delta + \epsilon_{r,c},$$

where $RegCovar_{rc}$ is the covariate enlisted at the beginning of each line. X is a set of regional controls that include the log of inhabitants without degrees, log of latitude, log of area, log of population density, and the log of the index of the quality of cultivable land from Ramankutty et al. (2002). Standard errors are clustered at the level of groups of regions (NUTS 1). Statistical significance is shown as follows: ***1%, **5%, *10%.

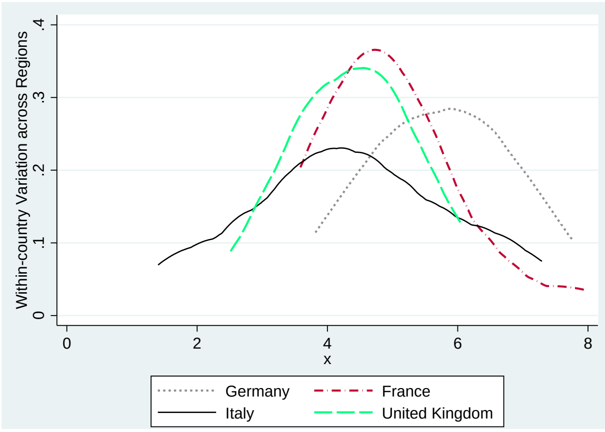
Online Appendix to

Innovation and Investment: The Role of Basic Education

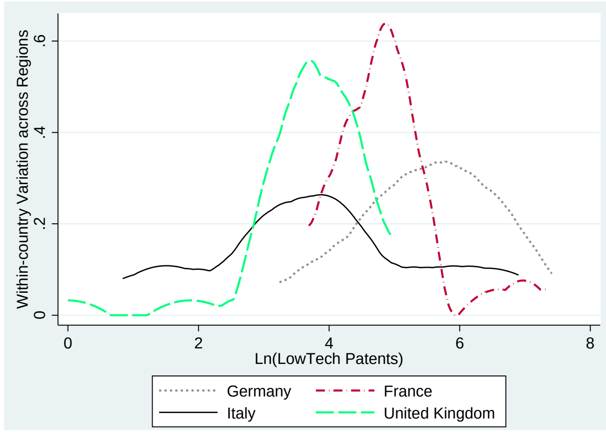
Francesco D'Acunto

Figure A.1: Distribution of Regional Variables Within Countries

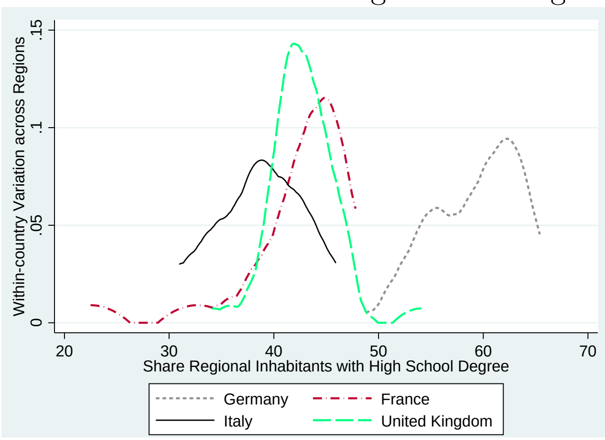
A. All Patents



B. Low-Tech Patents



C. Ratio Inhabitants High School Degree



D. Literacy Rates around 1880

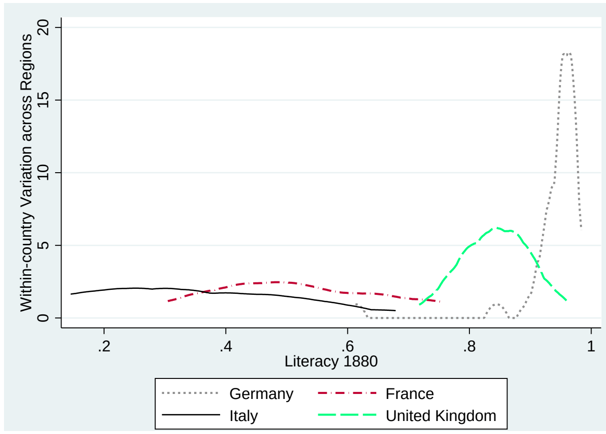
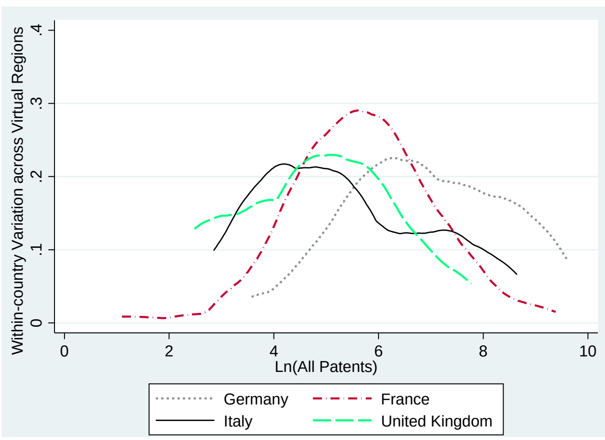


Figure A.2: Distribution of Virtual-Region Variables Within Countries

A. All Patents



B. Ratio Inhabitants High School Degree

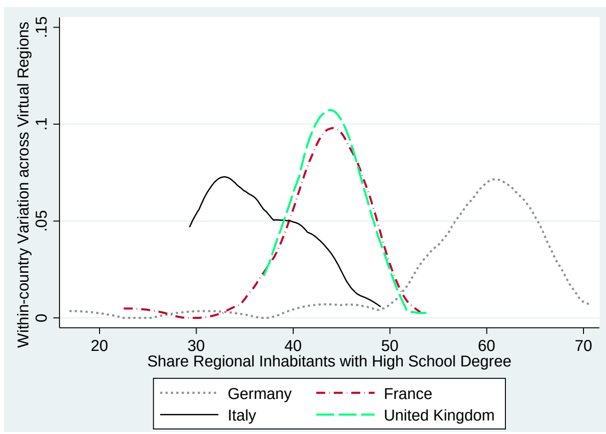


Figure A.3: Distribution of Firm-level Variables Within Countries

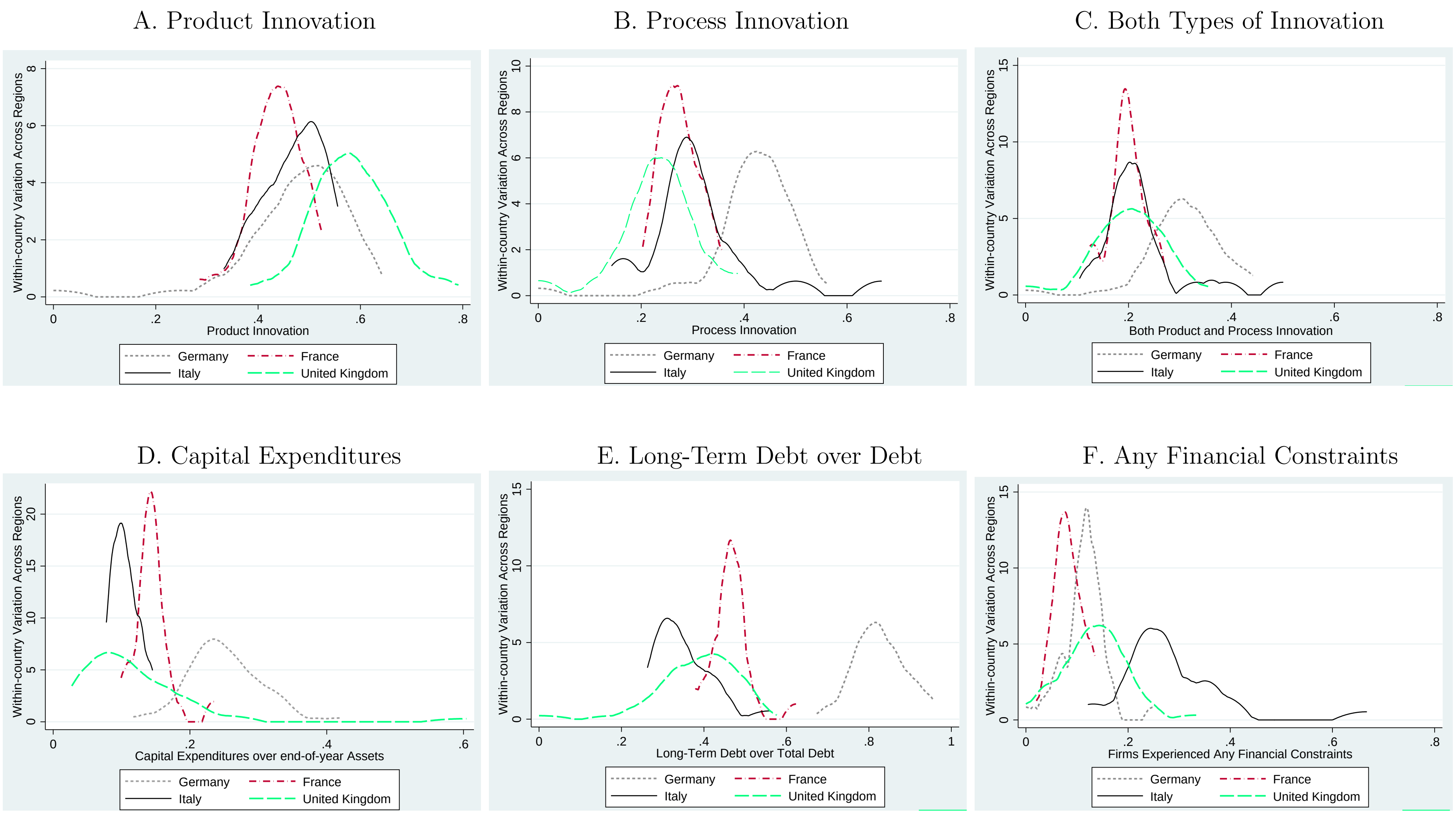
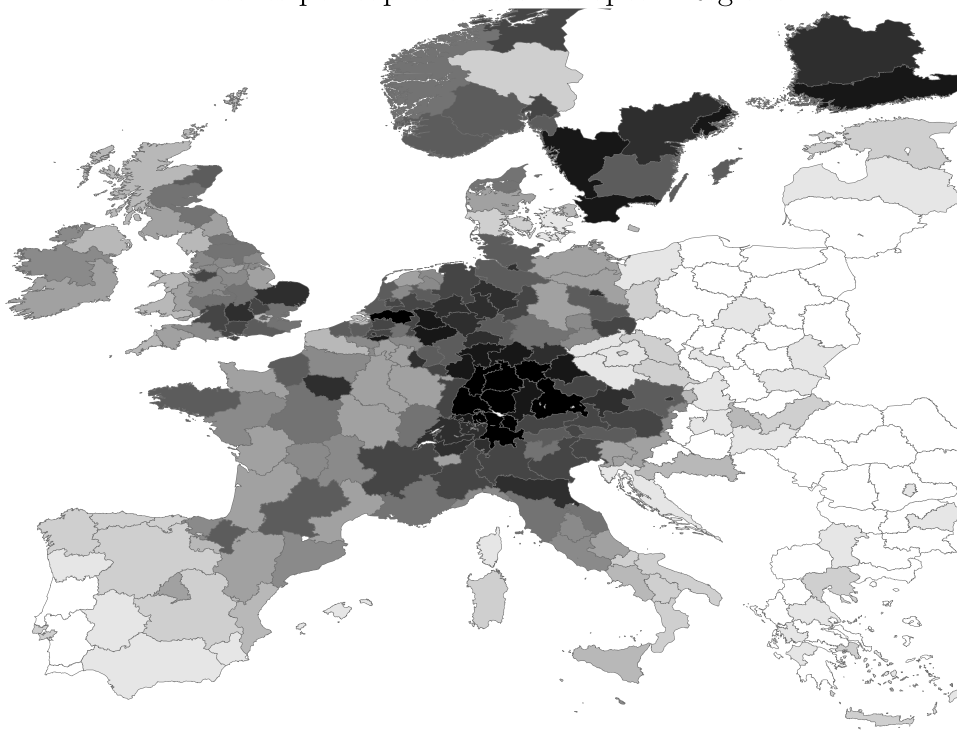
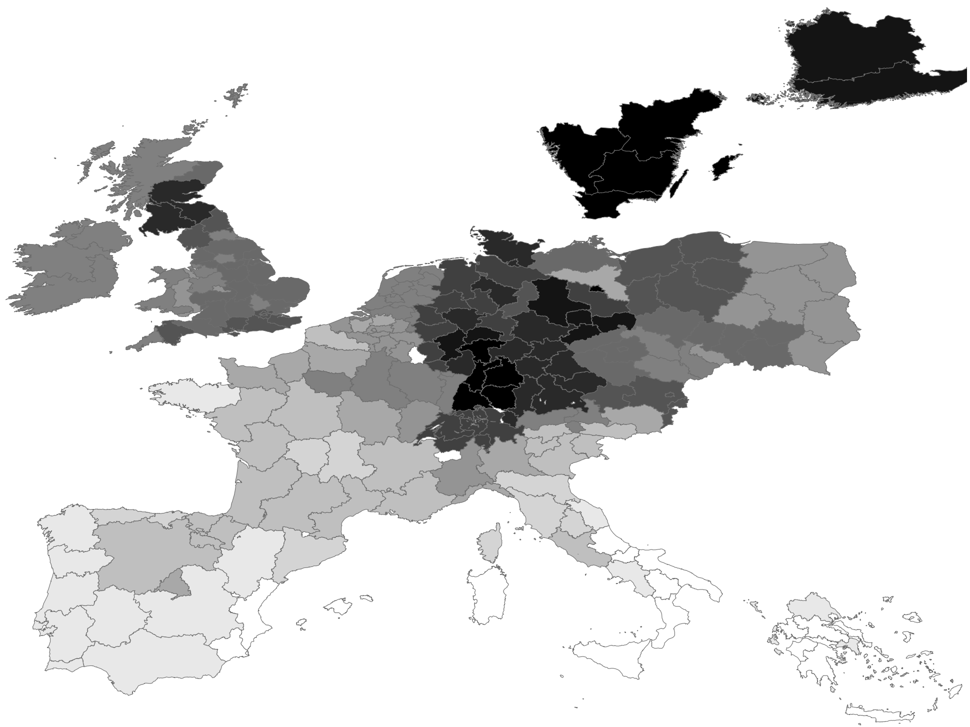


Figure A.4: Spatial Distribution of Patents per Capita and Historical Literacy

A. Patents per capita across European Regions



B. Literacy rates around 1880 across European Regions



Panel A plots the spatial distribution of the ratio of patents per capita for European Regions (NUTS 2) in 2005. The patent count is from the Patstat-Kites data set, based on the European Patent Office database. Panel B plots the spatial distribution of literacy rates in 1880. Literacy rates are from Tabellini (2010), and from primary sources for the regions of Greece and Spain. In both panels the darker a region is, the higher the value of the plotted variable.

Figure A.5: Commuting across Counties within Regions

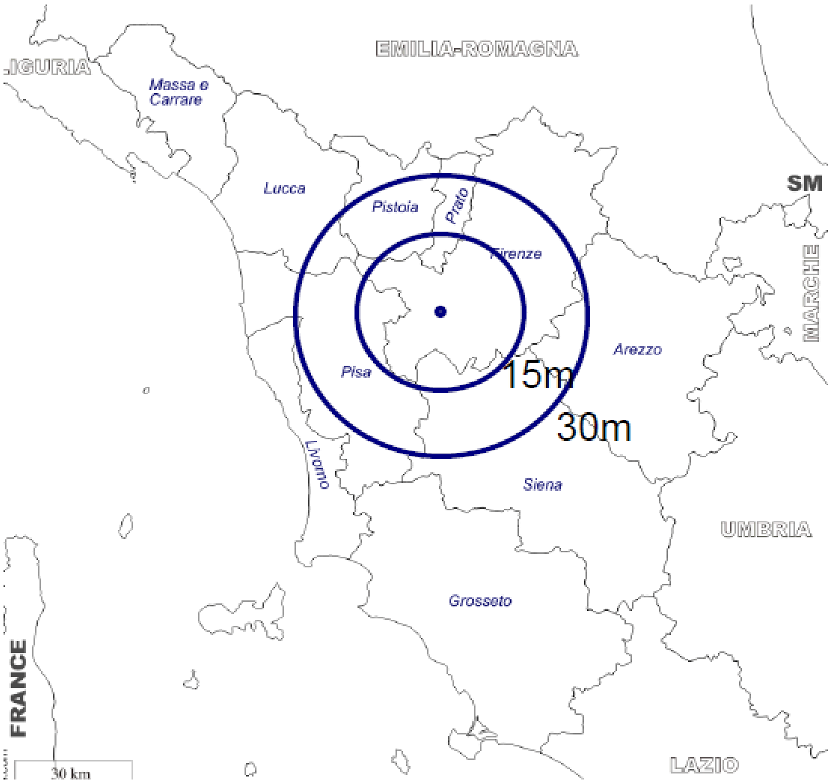
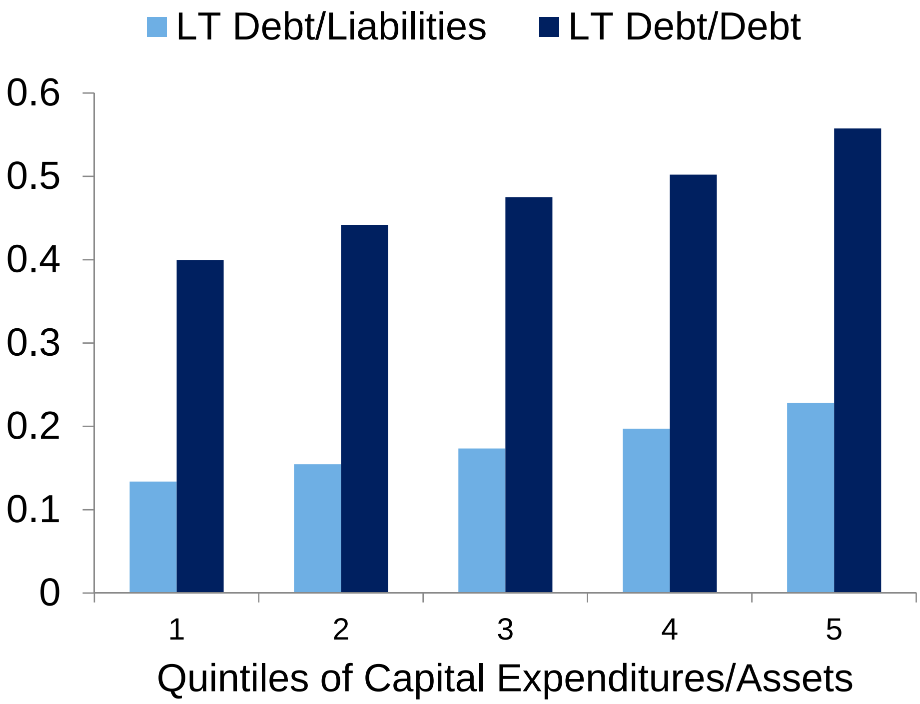


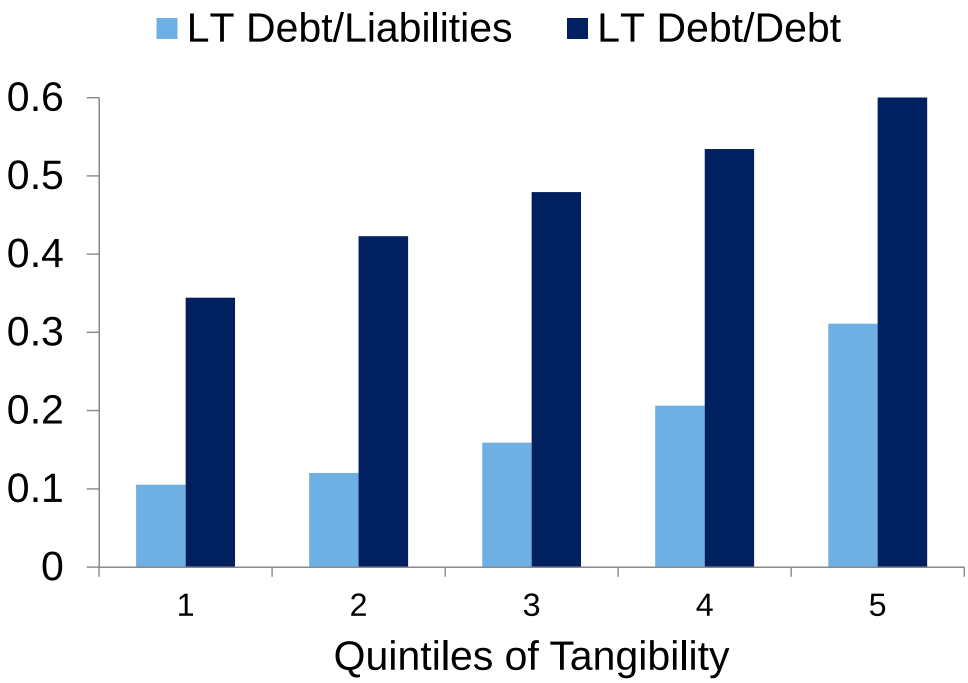
Figure A.5 illustrates the commuting times across counties of one of the 10% largest European regions, Tuscany. The blue point in the picture is the Empoli valley, in the Firenze county, an area that hosts several traditional manufacturing Small and Medium Enterprises (SMEs). Commutes as short as 15 minutes allow workers from 4 other counties to access the area: Pisa, Siena, Prato, and Pistoia. Commutes of about 30 minutes allow workers from 2 additional counties to access the area: Arezzo and Lucca.

Figure A.6: Long-Term Debt by Capital Expenditures and Tangibility

A. Long-Term Debt by quintiles of Capital Expenditures



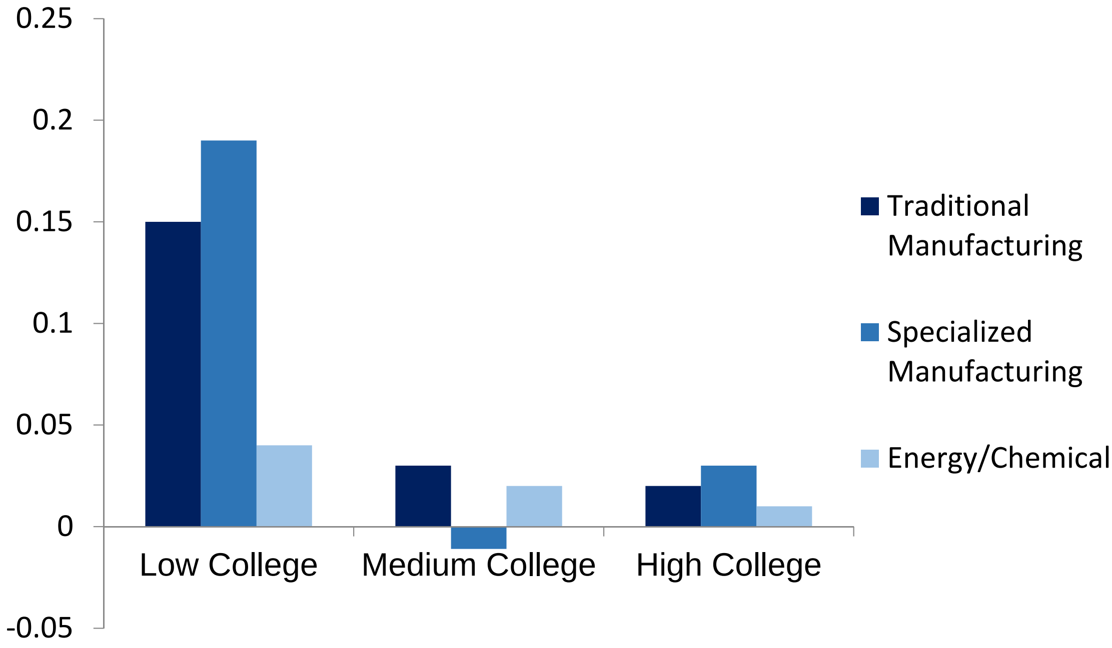
B. Long-Term Debt by quintiles of Tangibility



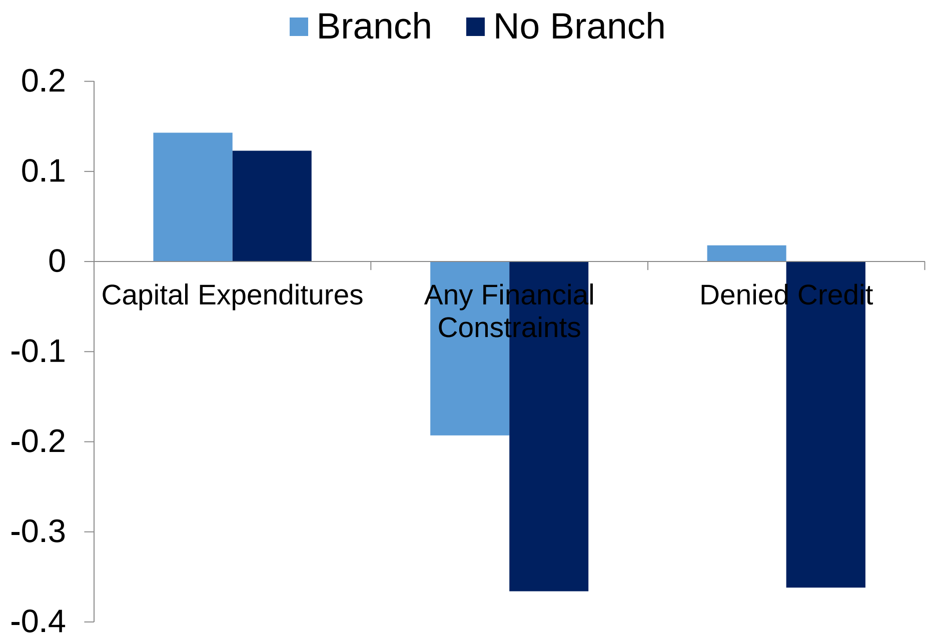
Panel A of Figure A.6 plots the average ratio of long-term debt over liabilities (light blue), and of long-term debt over total debt (dark blue) for the firms in the EU-EFIGE/Bruegel database across five equal-size groups of firms, sorted by their capital expenditures normalized by previous end-of-year total assets. Panel B of Figure A.6 plots the average ratio of long-term debt over liabilities (light blue), and of long-term debt over total debt (dark blue) for the firms in the EU-EFIGE/Bruegel database across five equal-size groups of firms, sorted by their tangibility, measured as the ratio between tangible assets and total assets.

Figure A.7: Additional Interaction Results

A. Basic Education and Product Innovations by Technological Intensity and College Employees

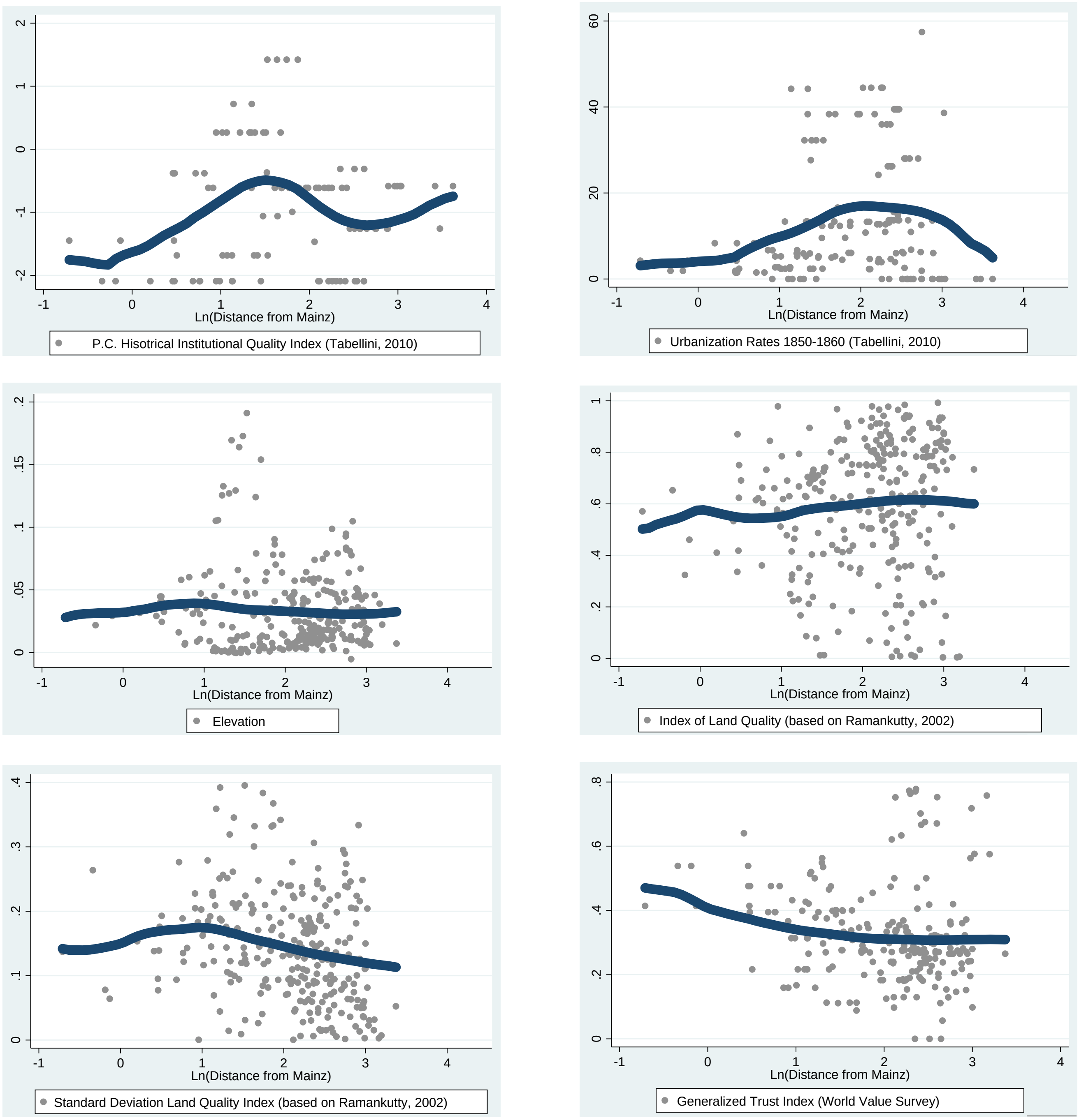


B. Basic Education and Investments, Financial Constraints by Branch Status



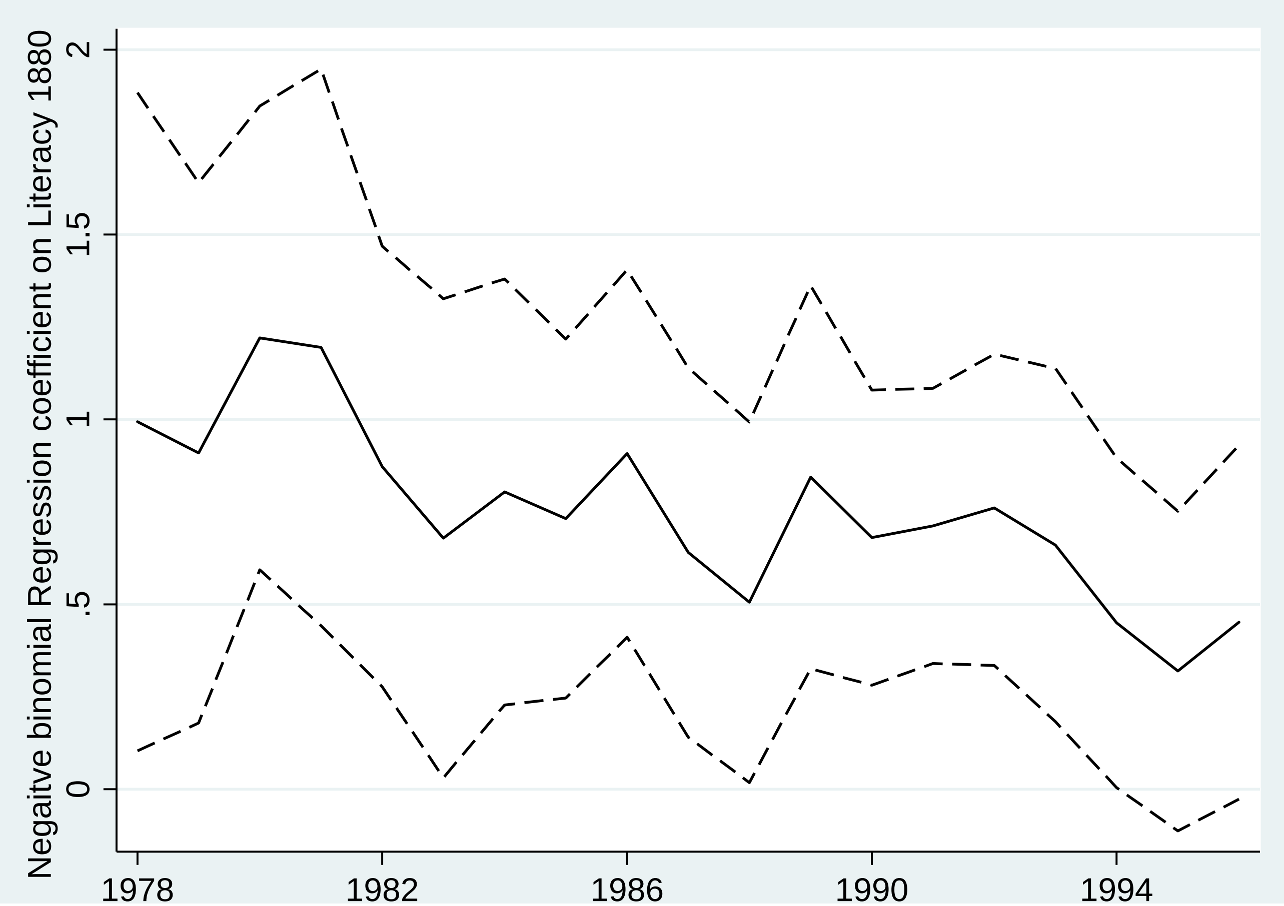
Panel A of Figure A.7 plots the estimated effect of regional high-school graduates on the likelihood that a firm in the EU/EFIGE-Bruegel data set engages in product innovation for firms sorted twice: across three equal-size groups based on the share of college-graduate employees in the firm, and across three Pavitt sectorial taxonomy groups based on the technological intensity of the sectors. Traditional manufacturing (dark blue) is the least technologically-intensive sector, whereas Energy/Chemical is the most technologically intensive sector. Panel B of Figure A.7 plots the effect of regional high-school graduates on the capital expenditures, the likelihood of financial constraints, and the likelihood of being denied credit, based on the survey responses of firms in the EU/EFIGE-Bruegel database. The effects are estimated separately for firms that are part of a business group, or branches (light blue), and firms that are autonomous (dark blue).

Figure A.8: Distance from Mainz and other Historical, Current Outcomes



The graphs in Figure A.8 plot the association between the dimension described in each graph and the log of the distance from Mainz across the European regions for which each dimension is observed. Dimensions include the Index of the quality of historical regional institutions from Tabellini (2010), the urbanization rate in a region in the period 1860-1880 from Tabellini (2010), the average elevation of the region, the index of quality of cultivable lands based on Ramankutty et al. (2002), the standard deviation of the indices of land quality for the cells that are included in each region by Ramankutty et al. (2002), and the index of generalized trust in a region computed as the regional average of the individual responses to the World Value Survey (Wave 9).

Figure A.9: Blue-collar Workers or Machines?



This graph reports the coefficient of the effect of historical literacy in a region as of 1880 on yearly regional patent counts from 1978 to 1996 estimated with negative binomial regressions of regional patents on the literacy rate in 1880 and a set of geographic and historical regional dimensions.

Table A.1: Historical Literacy and Current Regional and Firm-level Outcomes

A. Historical Literacy and Current Regional Innovation

	(1)	(2)	(3)	(4)
	All Patents		HighTech	Mid/Low Tech
Ln(Literacy 1880)	0.397***	0.232***	0.093	0.251***
	(0.080)	*(0.073)*	*(0.075)*	*(0.077)*
Country f.e.		X	X	X
Regional controls		X	X	X
Observations	228	227	222	222
N. of clusters	79	78	78	78
(Adjusted) R^2	0.184	0.888	0.832	0.864

B. Historical Literacy and Current Firm-level Outcomes

	(1)	(2)	(3)	(4)	(5)
		Innovation		Investment	Capital Structure
	Product	Process	Both	CapX	LT Debt/ Total Debt
Ln(Literacy 1880)	0.004	0.014*	0.017***	0.068***	0.056***
	(0.007)	*(0.007)*	*(0.006)*	*(0.017)*	*(0.008)*
Sector f.e.	X	X	X	X	X
Age f.e.	X	X	X	X	X
Size f.e.	X	X	X	X	X
Regional controls	X	X	X	X	X
Country f.e.	X	X	X	X	X
Observations	10137	10137	10137	10129	8080
N. of clusters	140	140	140	140	139
(Adj, Pseudo) R^2	0.066	0.023	0.028	0.118	0.246

Panel A of Table A.1 reports results for estimating OLS regressions of the log of regional patents on the log of regional literacy rates around 1880 and the set of controls in **??**, and limiting the variation within countries. Standard errors are clustered at the level of groups of regions (NUTS 1). Panel B of Table A.1 reports results for estimating the firm-level specifications of Table 11, where the log of regional high-school graduates is replaced with the log of the regional literacy rate around 1880. Standard errors are clustered at the level of regions (NUTS 2). In both Panels, statistical significance is shown as follows: ***1%, **5%, *10%.