Seeds of Knowledge: Premodern Scholarship, Academic Fields, and European Growth

Matthew Curtis[a] **and David de la Croix**[b,c]

[a]Syddansk Universitet; [b]IRES/LIDAM, UCLouvain, Belgium; [c]CEPR, London

Human capital is an engine of modern economic growth. Using a novel database of premodern European academics (1000–1800), we find that scholarship in the past also fostered growth. Combining secondary sources on the history of academia with data from worldwide library catalogs, our dataset measures both the quantity and productivity of scholars. Despite the fact that ideas and written knowledge was highly mobile, the birth places of scholars appear to have higher growth: we find that a 10% increase in scholarly output was associated with 1.4% higher income per capita in the region of the scholars' birth in 1900. Next, we use machine learning to group scholars into fields of specialization. Income per capita was positively associated with the share of scholars studying science (including mathematics, physics, and astronomy), botany, and one approach to theology, but negatively with the share studying law. Only the share of science and botany, however, seem to matter once we look within countries. We propose a mechanism consistent with this empirical evidence: scholars encourage their compatriots to accumulate human capital.

1. Introduction

The accumulation of knowledge is a crucial factor in economic development. It helps explain the West's prosperity and disparities in income among countries (1–3). However, measuring knowledge can be challenging, as it covers diverse subjects and can be embedded in various forms. In this study, we propose a novel method to quantify knowledge that developed in pre-industrial academia. We find evidence of a positive relationship between regional GDP per capita in 1900 and the birthplaces of academic scholars (university professors and members of academies of sciences and arts) during the period of 1000–1800.

Beyond measuring the quantity of academic knowledge in general, we also consider the types of knowledge produced. It has been argued that specific types of knowledge were important for economic growth. For example, scientific knowledge pushed the envelope of propositional knowledge, leading to future economic applications (4). Academic knowledge contributed to building better political and economic institutions as far back as the Middle Ages (5). Theologians promoted nuclear family structures (6) and held beliefs compatible with the spirit of capitalism (7), lawyers developed Roman and civil law encouraging trade (8), and physicians laid the ground for advances in botany (9).

To unravel which types of knowledge was more conducive to economic development, we first group scholars in different clusters using a machine learning algorithm, with each cluster representing one academic field. We find a particularly strong association between growth and the field related to mathematics and physical sciences and the field related to botany and life sciences.

Our dataset contains tens of thousands of scholars compiled from hundreds of secondary sources on the members of universities and academies. To measure the productivity of these scholars, we count every work and edition attributed to them in WorldCat. This approach is complementary to that of de Courson, Thouzeau, and Baumard (10), who use Wikipedia as both the index of individuals and the measure of output. The important difference is in which individuals are assigned a measure of productivity. Our sample is both more exclusive, in that it only considers members of academia, and be more inclusive, in that we are not selecting based on retroactive notability. Our paper is also complementary to that of Johnson et al. (11), who use texts as a measure of the local adoption of printing presses. We focus on the locations of authors, not publishers, and thus measure the human capital of scholars instead of the physical capital of printing presses.

Our fields of study are based on a list of subjects associated with the works by or about the author from the WorldCat Identities database. The subjects are based on the FAST subject terminology schema developed by OCLC (the organization that develops WorldCat) and the Library of Congress. Using these subjects, we use an unsupervised machine learning algorithm — k-means clustering — to assign each author to a cluster. This approach is similar in spirit to Grajzl and Murrell (12), Almelhem et al. (13), and Koschnick (14) who use machine learning to classify texts by topic. Their algorithms classify texts into topics. In our work, we classify scholars into fields based on their associated topics. We also share an interest in how types of knowledge matter for economic growth. Our approach is complementary, as we focus on the production of knowledge in academia and look at impacts over European regions.

While our scholars database ends in 1800, estimates of

Significance Statement

One theory for the rise of the West argues that universities and academics played a central role. However, there have be no quantitative studies of historical academia and growth for Europe as a whole. This paper develops a methodology to measure academic productivity using a large novel database of scholars 1000–1800. We find that the output of academics predicts 19th century economic growth, providing Europe-wide evidence that the sciences paved the way for the Industrial Revolution. Moreover, approaches to theology and legal systems also mattered for economic development. Despite the fact that ideas and written knowledge were highly mobile, the birth places of scholars in particular appear to have higher growth, suggesting scholars played a role-model for the next generations.

income are too sparse before 1900 to compare all regions where scholars were born. Thus we focus on outcomes in 1900, using estimates of GDP per capita for the contemporary NUTS2 regions from the Rosés-Wolf database on regional GDP (15). One advantage of this approach is we are looking at gains from both the First and Second Industrial Revolutions. While during the First Industrial Revolution there was a major role for the human capital and idiosyncrasies of craftsmen and inventors, the Second Industrial Revolution saw a more direct pipeline between scientific knowledge, applied innovations, and an educated workforce (16).

2. Scholars, Universities and Academies

Medieval universities concentrated on four main fields: theology, law, arts and humanities, and medicine. Their impact on society is well described by Pedersen (17). "The faculty of arts gave a basic education to grammar school boys, many of whom would become teachers themselves and contribute to the increase in literacy of the population at large. Others would go on to one of the higher faculties to prepare themselves for other professions. The faculty of medicine produced medical practitioners; the faculty of laws created future administrators with expert knowledge in canon or civil law, and the faculty of theology provided teachers for the episcopal schools, were the ordinary parish priests were educated." Academies were usually created later, in the 17th-18th century, responding to a need of developing new fields of research which were not traditionally taught at universities. The academies range from clubs of amateur naturalists or local historians to eminent societies, gathering the best scholars, publishing journals, and building a network of corresponding members, called the Republic of Letters (2, 18).

The full database of scholars contains information on 60,001 scholars who were appointed to universities or were nominated to academies over the period 1000–1800. The data were harvested manually from 535 different secondary sources on the history of universities and academies. We took the list of universities from (19) and the list of academies from (18), and added to this the language academies, the most important Italian Renaissance academies from (20), and several other higher education institutions which conferred academic degrees. More information on the criteria to include scholars in the database is available from (21), while some global statistics are provided in (22) and in the various issues of the *Repertorium Eruditorum Totius Europae*.

To assign a measure of productivity to each scholar, we use the Worldcat search engine which provides references to the collections of thousands of libraries around the world. We count the number of "works," i.e. publications by the author. This measure thus cover both output of the scholar and impact. Worldcat provides a good approximation of the population of known European authors, for example, Chaney (23) compares the Universal Short Title Catalogue (24) to the references in the Virtual International Authority File (VIAF), on which WorldCat is based. Chaney successfully locates 81% of USTC authors in the VIAF. hence scholars with missing Worldcat publications were likely unproductive.

3. Identifying Academic Fields

For each scholar with a Worldcat reference (excluding the persons who are honorary members of academies), we collected the tag cloud of their "Associated Subjects." We drop subjects associated with fewer than 30 scholars or that are about a specific country (e.g. "French history"). This leaves us with 1,360 subjects and 16,149 scholars with at least one subject. We partition the data into k clusters, minimizing the total within-cluster sum of squared deviations. The choice of k can be made using various criteria. We minimize the Bayesian information criterion (BIC) and determine that ten clusters is the most informative yet parsimonious way to describe academic fields. Table 1 presents the ten clusters. The first column contains a description we chose to represent the various subjects included in the cluster. Column 2 gives the total number of published scholars in each cluster. One cluster is much bigger than the others; it appears to contain both classicists and scholars who were unrelated to any other cluster. The smallest cluster is Botany, with 543 persons.

To better grasp the nature of each cluster, we show in Column 3 the names of the scholars belonging to the cluster who published the most. Column 4 gives the median number of publications of scholars in each cluster. Theology 2 leads and Classics lags. Column 5 shows the date of activity of the earliest scholar in each cluster. It shows that all ten clusters started before 1200, thus having deep roots in the Middle Ages. The last column shows the median year of activity in the cluster. Law is the cluster with the earliest median date, while Politics is the cluster with the most recent median date.

The clusters are further explored in the Appendix. The most important topics and scholars by cluster are described in Appendix B. Most clusters are strongly associated with a few key terms, however the Classics cluster is not. Classics contains authors who write on many diverse topics, perhaps related to the Humanistic Revolution. Appendix C plots the shares of scholars by cluster over time. In Appendix D, we provide ten graphics with names of published scholars over time by cluster, allowing to see through whom each field has medieval roots.

Theology is the only field to have two clusters (see maps in Appendix H). The division between Theology 1 and Theology 2 is related to the Catholic-Protestant divide, but is not a simple denominational split. In Theology 1, we find some leading figures of Catholicism such as Aquinas (professor at University of Paris 1252–72 and Naples 1272–4), Bossuet (member of Académie Française 1671–1704), and Robert Bellarmin (professor at the Gregorian University in Rome 1576–1593) but also some unorthodox catholics such as Pascal (member the Mersenne academy of c. 1639, close to Jansenism, a controversial Catholic movement with similarities to Calvinism) and some important Protestant figures such as Gilbert Burnet (professor of Divinity at the University of Glasgow 1669–74, and member of the Royal Society). Theology 2 is led by the main figures of Protestantism, such as Luther (professor at University of Wittenberg 1508–46), Melanchthon (professor at University of Tübingen 1512–18 and Wittenberg 1518–60), John Wesley (fellow of Lincoln College at University of Oxford 1725–7), and Jean Calvin (professor at the University of Geneva 1541–64). But it also includes medieval (Catholic) theologians such as Hugues de Saint-Victor (University of Paris 1133–41). Looking at the subjects with the highest frequency in both clusters, we find "Catholic Church" and "Clergy" in Theology 1, and "Bible" in Theology 2.

Scientific fields are split in three clusters: Sciences,

Table 1. Clusters of WorldCat Topics

Cluster / Field	N. Scholars	Top 3 Names	Median N. Publ.	Earliest Year	Median Year
Theology 1	1581	Aquinas, Bossuet, Pascal	143	975	1615
Theology 2	940	Luther, Melanchthon, Wesley	315	1039	1671
Politics	990	Swift, Machiavelli, Corneille	184	1043	1756
Law	727	Stryk, Bentham, Bohmer	156	1090	1593
Science	661	Newton, Euler, Galilei	177	1116	1714
Classics	7317	Schiller, Erasmus, Pope	54	970	1712
Philosophy	653	Rousseau, Kant, Diderot	258	980	1700
Botany	543	Linnaeus, Bernardin, Trew	189	1176	1753
Culture	1086	Arouet, Humboldt, Homman	211	1140	1749
Medicine	1651	Haller, Hohenheim, Gessner	125	1025	1698

Note: Clusters estimated by k-means clustering. Top 3 Names are the top three scholars assigned to a cluster based on their number of publications.

with the subjects "Mathematics", "Astronomy", "Geometry", "Physics", led by Newton (professor at University of Cambridge 1661–1696, member of several academies), Euler (professor at University of St Petersburg 1727–41, member of several academies), and Galilei (professor at University of Pisa 1589–92 and Padua 1592–1610). The cluster Botany includes the subjects " Plants" and "Natural History", and led by Linnaeus (professor at University of Uppsala 1742–78, and member of many academies). The cluster Medicine includes subjects "Human anatomy" and "Surgery". Together with the clusters on Politics, Law, and Philosophy, the clustering procedures seems to lead to a very coherent set of academic fields. Only Classics and Culture have vague boundaries. We are thus confident interpreting these clusters as academic fields.

4. Academic knowledge and regional development

We now analyze whether academic knowledge is associated with historical development at the subnational level. This allows us to determine if scholarship matters both at a local and a national level. We interpret a higher GDP per capita in 1900 as evidence of economic growth. Before 1800, GDP per capita was restricted by the Malthusian trap, albeit with some geographic and temporal variation. Moreover, we control for initial conditions with log total urban population in 1800 from (25), log ruggedness from (26), and log area of the region in km^2. Together, these controls are proxies for the economic development of the region c. 1800. Controlling for urban city population is particularly important as historical urban population levels are a measure of the size of the market for texts (11). With these controls, and given the low initial levels of development, we interpret a higher GDP in 1900 as evidence of stronger 19th century economic growth.

Figure 1 shows the geographical area we cover with the NUTS2 regions. The map's background color for each region reflects its GDP per capita in 1900, with darker shades indicating higher levels. Color dots indicate the place of birth of scholars belonging to two example fields. Red dots correspond with scholars belonging to the field of Law, blue dots with scholars belong to the field of Science. In a way the map captures the stylized facts we know about the Scientific Revolution with its main centres initially in Northern Italy, then the Netherlands, and after that in Paris and England.

When summing over scholars at the regional level, we weight each scholar by a function of the number of his publications. The number of publications, which includes multiple editions and translations, ranges from 1 to 111,660 (Martin Luther). It is not reasonable to assume that Luther worth a hundred thousand obscure theologians (those with only one publication). If, instead of the number of publications, we take its square root, Luther would be worth 334 obscure theologians. If we take the fourth root of the number of publications, Luther would be worth 18 obscure theologians. Galileo would be worth 9 mathematicians with one publication. We adopt this last formula, which gives a weight from 1 to 18 to each scholar. (In Appendix E, we show that this choice of weighting does not drive our results.)

We estimate the following regression model:

$$y_{r,s} = \alpha_0 + \alpha_1 \log(n_{r,s}) + \sum_{c=1}^{10} \beta_c share^c_{r,s} + \beta X_{r,s} + \phi_s + \epsilon_{r,s} \quad [1]$$

where $y_{r,s}$ is the GDP per capita in 1800 for region r of country s; $n_{r,s}$ is the weighted sum of published scholars born in r from 1000–1800, weighted by their number of publications raised to the power of 0.25; c is one of the ten fields identified by the K-means algorithm; $share^c_{r,s}$ is the share of $n_{r,s}$ that belong to field c; $X_{r,s}$ is a vector of controls, ϕ_s is a country fixed effect, and $\varepsilon_{r,s}$ is an error term.

The control variables include log total urban population in 1800 from (25), log ruggedness from (26), log area of the region in km^2, and, in some specifications, country fixed effects.

In our main set of regressions, scholars are allocation to their region of birth. We estimate a second set in which scholars are allocated to their region of activity (see Appendix G). We find that the first set gives stronger results. This suggests the presence of mechanisms beyond the mere effect of universities and academies on the region in which they are located. Moreover, in Appendix F, we show that these results robust to the inclusion of controls for the location of such academic institutions.

Figure 2 presents the results (see Appendix A for the results in table form). As shown in the first line, we find an overall association between $\log(n_{r,s})$ before 1800 (the weighted sum of published scholars) and GDP per capita in 1900. A one percent increase in the weighted sum of scholars born between 1000 and 1800 in region i is associated with a 0.14 percent increase in GDP per capita in 1900, all else equal. The estimate is reduced in magnitude but still significant after adding country

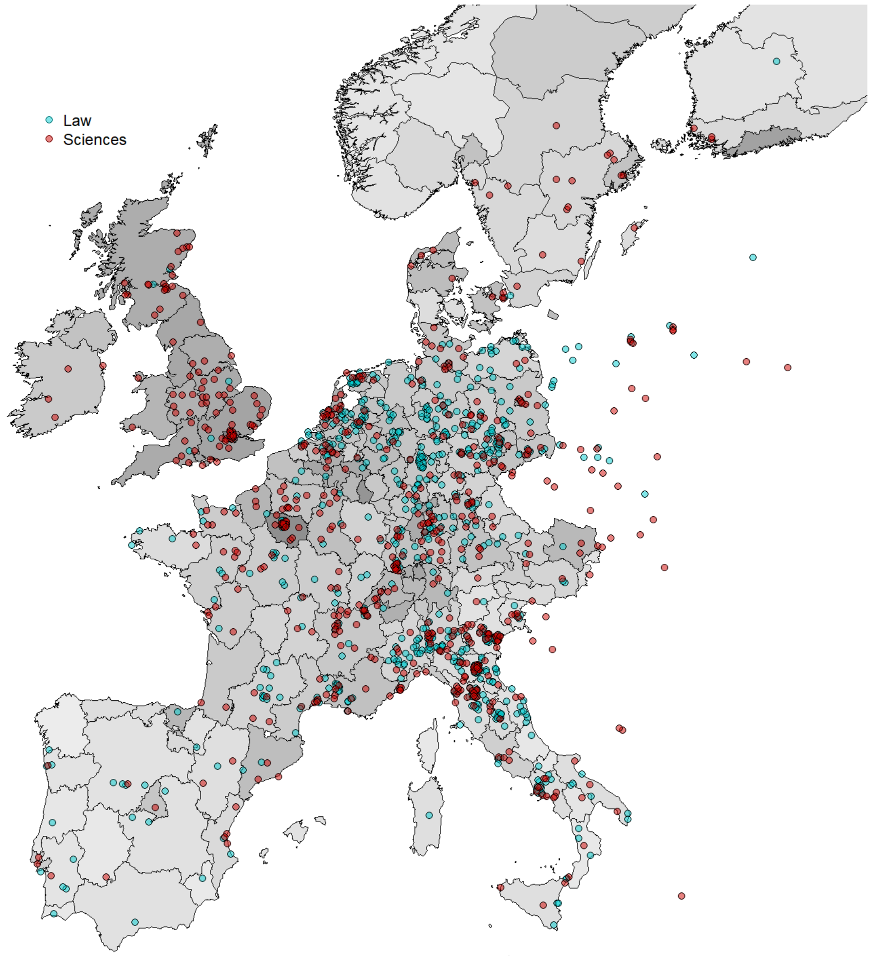

Fig. 1. Map of Birth Place of Law and Science Scholars

Note: Every scholar is assigned a field and a birth NUTS 2 region. GDP per capita from the Rosés-Wolf database on regional GDP (15).

fixed effects (second line). This shows that human capital in the past is associated with future growth. Below, we argue that this is likely a causal effect. Regardless of the exact mechanism, our findings lend credence to theoretical frameworks in which human capital plays a role in development.

When we additionally look at the shares of the different fields (Figure 3), we find that the fields Theology 2, Science, and Botany have a positive association with growth. The field Law has a negative association. In that regression, the reference category is the share of scholar in Classics. All else equal, a ten percentage point increase in the number of scholars that are in the field Theology 2 (at the expense of Classics) is associated with a 8.9% increase in GDP per capita in 1900. For Law, Science, and Botany, the changes associated with a 10 percentage point increase are −7.5%, 11.4%, and 8.2%.

We also still estimate the impact of the total number of scholars (third line of Figure 2), with a 1 percent increase in the weighted sum of scholars born between 1000 and 1800 in region i being associated with on average a 0.10 percent increase in GDP per capita in 1900, *ceteris paribus*. Therefore, the coefficients for the shares are estimating the *additional* impact from specialization in a field compared to the others. Scholarship, regardless of field, is associated with higher GDP per capita.

Theology 2 and Law appear to vary substantially across countries (using contemporary boundaries). Theology 2 is more common in Protestant countries, and Law is rare in common law Britain (Figure 1). As Protestantism and common law are commonly studied as determinants of growth, the associations we find might be related to more broad factors relating to religious sects and legal systems (7, 27). To control for any country-specific characteristics, we add country fixed effects (again, using contemporary boundaries). The association disappears for Theology 2. For Law, Science and Botany, the changes associated with a 10 percentage point increase are −3.1% 8.5%, and 5.9%. This suggests that Theology 2, and partially Law, are related to growth through some mechanism occurring at the national level.

In Appendix A we also look at GDP per capita in 2015 (from Eurostat), the positive overall association remains. A one percent increase in the weighted sum of scholars born between 1000 and 1800 in region i is associated with a 0.8

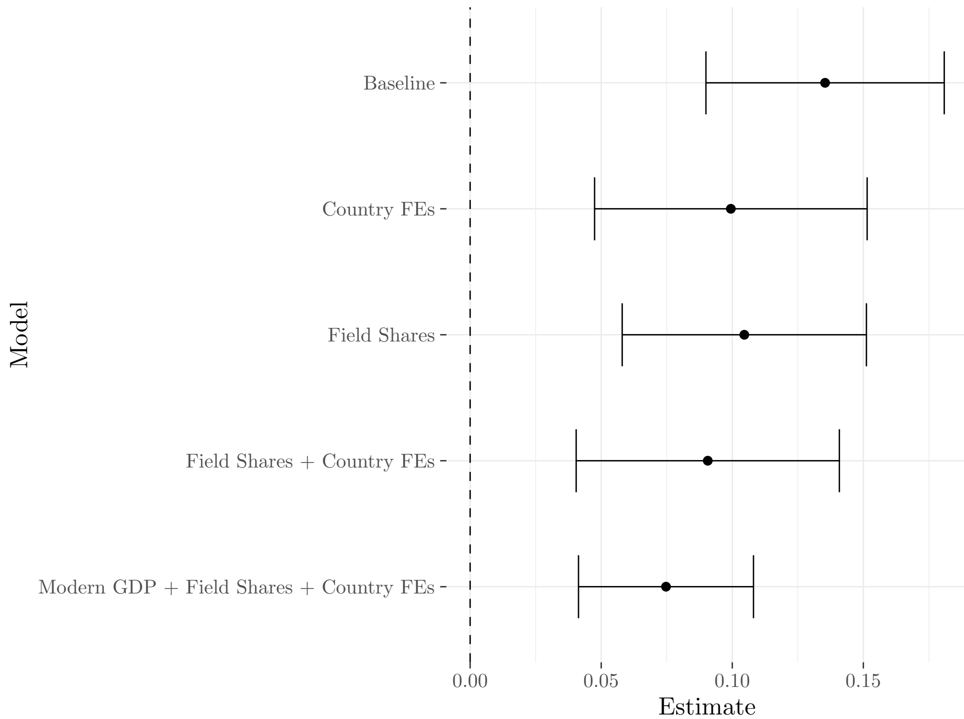

Fig. 2. Regional GDP per capita and academic output, 1000–1800

Note: 95% confidence intervals displayed. The unit of observation is a NUTS2 region. Every scholar is assigned a weight equal the fourth root of their number of publications, and to a birth region. GDP per capita from the Rosés-Wolf database on regional GDP (15). $\log(n_{r,s})$ is the log of the weighted total of scholars plus one. All regressions include as controls the log total urban population in 1800 from (25), log ruggedness from (26), and the log of the region's area.

percent increase in GDP per capita in 2015. Having a high concentration of scholars gave regions an initial advantage, and while the advantage is smaller in 2015, the other regions have not fully converged. On the other hand, the initial advantage from having a high proportion of scientists and botanists appears to have fully converged by 2015. Perhaps scientists were particularly important for the early adoption of the technologies of the Industrial Revolution, leading to an initial but temporary edge.

5. Mechanisms

Given the high mobility of ideas and written knowledge, it is surprising to find a local effect of the density and quality of scholars in the past. We investigate whether our results could be spurious and what could be the mechanisms behind them.

One concern with these results might be that some omitted variable determines both the number of scholars born and future GDP per capita. In particular, something might increase the local demand for or supply of scholars. To address these concerns, we identify group of scholars who potentially had less influence on their region of birth. If these weakly attached scholars have minimal effect on their region of birth, then it is unlikely that there is an omitted variable increasing both the number of scholars and GDP.

The first group are scholars who died in a foreign country. This is a proxy for scholars who emigrated, and thus had a weaker connection to their home region. The second group are those who died after the Peace of Augsburg (1555) in a country with a different state religion than their home country (omitting Germany due to its religious heterogeneity). This proxies for what we term émigrés, that is scholars who migrate due to a religious or political conflict in their home regions. The most prolific of these émigrés was René Descartes, who was born in France and died in Sweden. He did not convert to Protestantism, but was placed on the *Index Librorum Prohibitorum* in 1663. Other notable examples are Helen Maria Williams, an English Girondin revolutionary, Alban Butler, an English Catholic priest, and Jacques Abbadie, a French Anglican minister. The third group has a more inclusive definition of migrant, consisting of any scholars who died in a different location from their birth. The final group are scholars who died before age 40. These scholars had less time to build a local reputation even if they had produced scholarly works.

The regressions are of the form:

$$y_{r,s} = \alpha_0 + \alpha_1 \log(n_{r,s,i}) + \log(n_{r,s,j}) + \beta X_{r,s} + \phi_s + \varepsilon_{r,s} \quad [2]$$

Notation is the same as in Equation (1). $n_{r,s,i}$ is the sum of published scholars in the group i of interest, born in r from 1000–1800, weighted by their number of publications raised to the power of 0.25; is $n_{r,s,j}$ the same but for scholars not in the group of interest. We also run the same regressions for regions of death.

As shown in Figure 4 Lines 1–10, the output of those scholars weakly attached to their birthplace appear to have little to no association with growth after controlling for the output of the rest (see Appendix A for the results in table

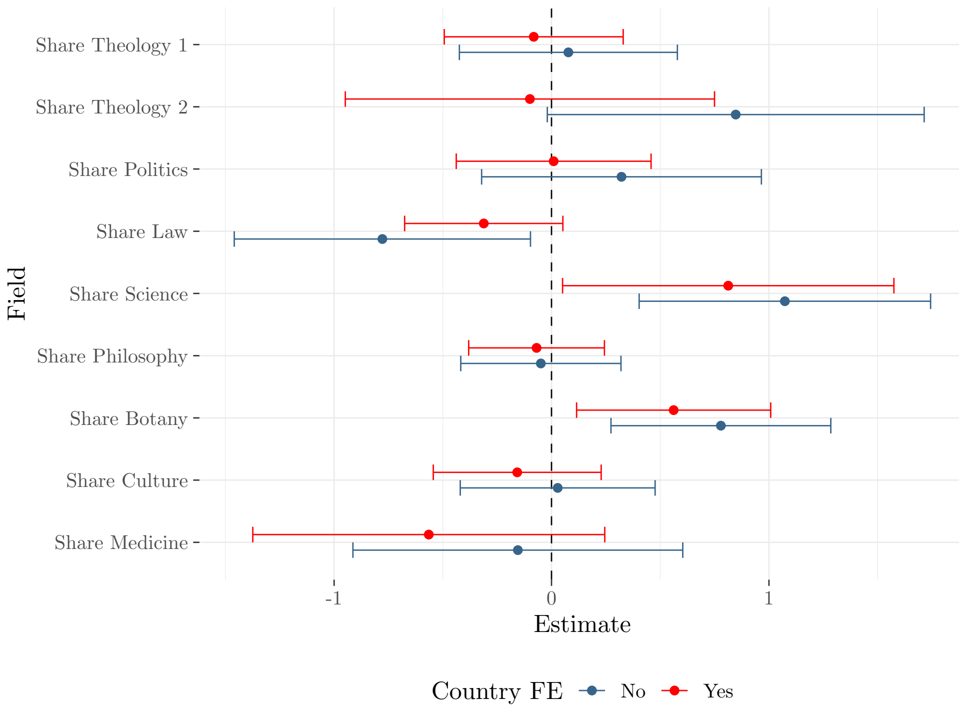

Fig. 3. Regional GDP per capita and academic fields, 1000–1800

Note: 95% confidence intervals displayed. The unit of observation is a NUTS2 region. Every scholar is assigned a field, a weight equal the fourth root of their number of publications, and to a birth region. Shares are the share of the total scholars who are assigned to a given field. GDP per capita from the Rosés-Wolf database on regional GDP (15). All regressions include as controls the log of the weighted total of scholars, log total urban population in 1800 from (25), log ruggedness from (26), and the log of the region's area.

form). Moreover, the coefficients for the output of scholars who were not weakly attached are very similar to the coefficient of the baseline regression. This suggests that the associations in Table 2 are driven by scholars with a close association with their birthplaces.

A contrasting result is found in the last two lines of Figure 4, which splits scholars born before and after 1600. This crude periodization attempts to split the sample roughly before and after the Scientific Revolution. A scholar born in 1600 could read Bacon's *Novum Organum* at age 20 and Galilei's *Dialogue Concerning the Two Chief World System* at age 38. While the coefficient for the earlier scholars is slightly weaker, the difference is marginal. In other words, the output of pre-Scientific Revolution scholars seem as important as that of post-Scientific Revolution scholars.

These results suggest that there really was a mechanism tying scholars to growth in their place of birth. One possibility is that successful scholars encouraged others from the same region to accumulate human capital. Early Modern Europe's "Republic of Letters" was a small elite network, but provided notable scholars with both prestige and financial patronage and was relatively open to new talent (28). Examples of scholars born nearby may have been illustrative of the potential returns to human capital.

The available data for regional human capital for Europe in 1900 are more limited than those for income. Nevertheless, Figure 5 estimates the same regressions as Figure 2 for a measure of human capital: numeracy (see Appendix A for the results in table form). To measure numeracy, we use the ABCC index from Baten and Hippe (29). This index is defined as $125(1-s)$, where s is the share of reported ages between 23 and 72 which end in 0 or 5. It measures a very rudimentary level of human capital: do people know their own age. We find that, even controlling for GDP in 1900, areas with a greater number of scholars born had higher human capital in 1900. Depending on the specification, 10 percent increase in the weighted sum of scholars born between 1000 and 1800 in region i is associated with an increase in the ABCC index of 5.0–11.5. This corresponds to a 5.0 to 9.2 percentage point decrease in people who round their reported age to 0 or 5. In other words, areas with higher scholarship had higher *lower-tail* human capital, even after controlling for economic development.

6. Discussion

We find a strong relationship between economic growth and pre-modern European scholarship. Our findings support the view that upper tail human capital was important for growth (30). Moreover, we find that certain fields of scholarship had a stronger influence on growth than others.

Perhaps it is not surprising that we find that Science and Botany were particularly important. Fundamental scientific research paved the way for future applied technologies. For example, engineering has been critical to the development of infrastructure and technology throughout history (31). Engineering was not part of curricula in the period we consider

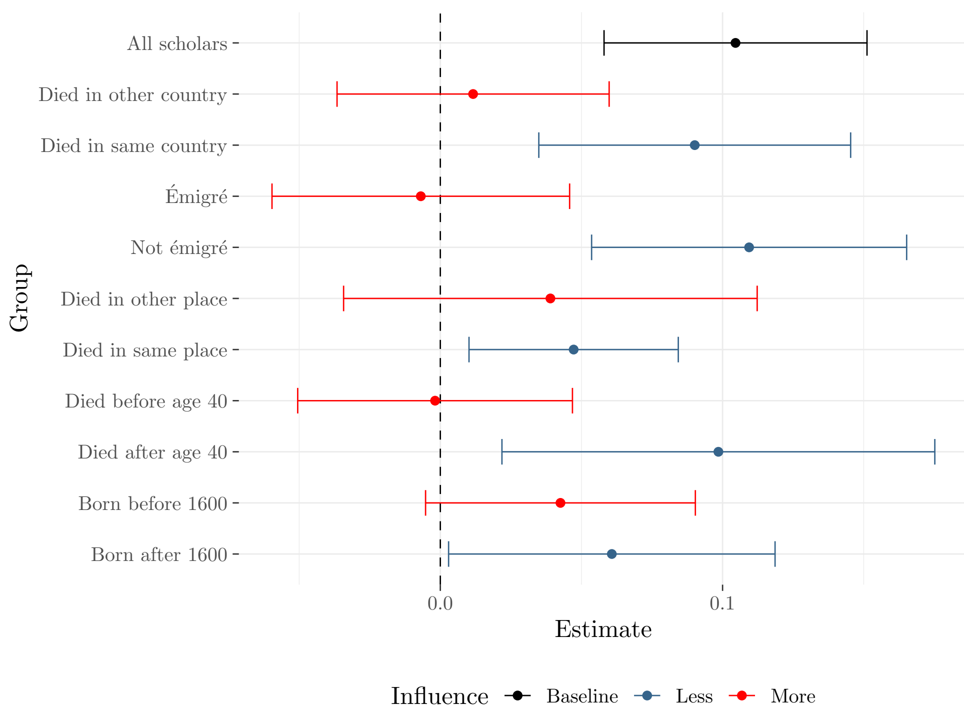

Fig. 4. Analysis by strength of attachment to birth region

Note: 95% confidence intervals displayed. The unit of observation is a NUTS2 region. Every scholar is assigned a weight equal the fourth root of their number of publications and to a birth region. GDP per capita from the Rosés-Wolf database on regional GDP (15). $\log(n_{r,s,i})$ is the weighted total of scholars belonging to a subset of scholars. All regressions include as controls the log total urban population in 1800 from (25), log ruggedness from (26), and the log of the region's area.

(1000–1800), but is strongly grounded in mathematics and physics, two important components of our field Science. Medical research and advancements have been crucial to improving public health, curing diseases, and extending lifespan, in particular in the nineteenth century. Modern medicine is based on natural sciences such as botany, which appears as a strong correlates of growth as well.

The positive role of a certain type of Theology and the negative role of Law are also interesting. Both appear only when comparing across, not within, countries. Theology 1 emphasizes subjects such as the clergy, whereas Theology 2 subjects such as the Bible. Theology 2 is more closely related to Protestantism, although not exclusively. Countries with a higher share of this type of theology are richer at the end of the nineteenth century. According to Max Weber (7), Protestantism facilitated the rise of capitalism by instilling a set of values and attitudes that were conducive to economic development. Perhaps the scholars in the field Theology 2 were fostering these cultural norms. Cantoni et al. (32) find that religious competition in Protestant German states led to graduates seeking non-religious employment after the Reformation. Perhaps the students of scholars in Theology 2 were more likely to apply their human capital in secular activities. This interpretation would be particularly compelling if Theology 2 is capturing some measure of religious competition; this could explain its inclusion of unorthodox Catholic scholars as well as Protestant theologians. However, these associations only show up in regressions without country fixed effects, suggesting that they are related to a process that occurs at the national level, and not at the region of a scholar's birth.

It is more difficult to interpret the negative role of Law. It could be that the share of law among academic scholars reflects the local legal system. Indeed, in common law countries, legal education and training are often not solely confined to universities, and there is more emphasis on practical training through apprenticeships, clerkships, and other forms of legal practice. Civil law countries have more lawyers in academia, and there is a large literature showing that these countries tend to perform less well than common law countries (27). While the association remains after controlling for country fixed effects, it is substantially diminished, again suggesting this is mostly a process occurring at the national level.

We believe that the mechanism that most likely explains our results is that scholars inspire compatriots to accumulate human capital. This inspiration could be through social networks. Leonhard Euler was born in 1707 in Basel as the son of Paul Euler, a Reformed pastor. As a college student at the University of Basel, Paul had befriended Jacob and Johann Bernoulli (1655 and 1667, Basel). Johann later convinced him to let his son Euler study mathematics instead of theology. Both Bernoullis, notable mathematicians in their own right (they are the 82nd and 83rd most prolific members of the field Science), thus directly contributed to young Euler becoming the most second most productive member of our field Science (behind only Isaac Newton).

This inspiration could also be an indirect effect on future

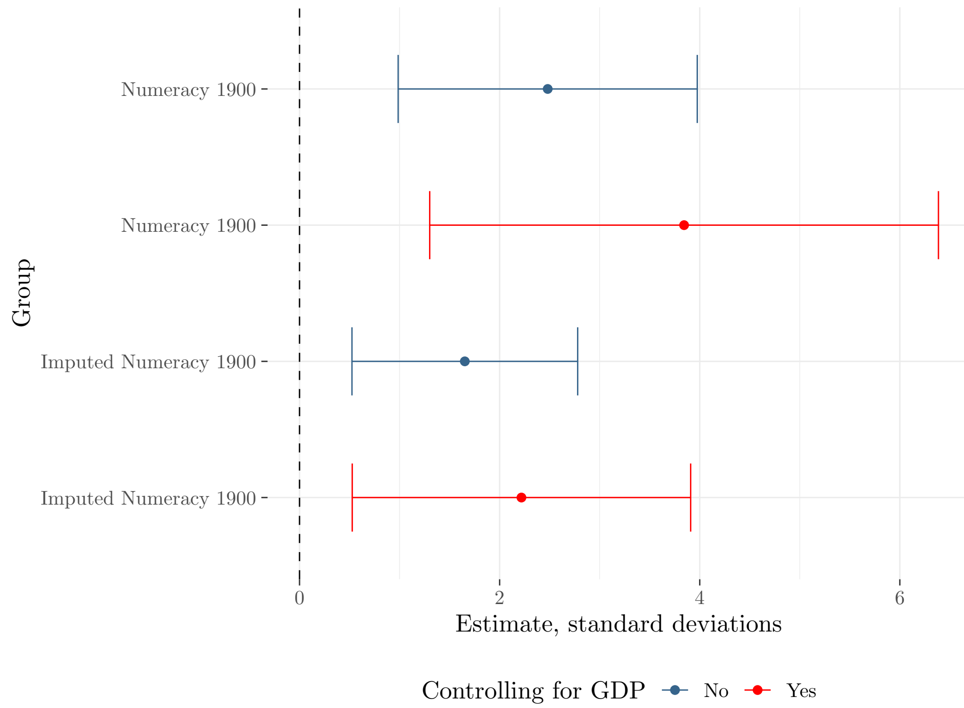

Fig. 5. Human capital and academic fields, 1000–1800

Note: 95% confidence intervals displayed. x-axis normalized to be standard deviations. The unit of observation is a NUTS2 region. Every scholar is assigned a weight equal the fourth root of their number of publications, and to a birth region. Numeracy is the ABCC index from Baten and Hippe (29): $125(1-s)$, where s is the share of reported ages ending in 0 or 5. Imputed numeracy assigns a value of 1 to Germany and Scandinavia in 1900. GDP per capita from the Rosés-Wolf database on regional GDP (15). All regressions include as controls fixed effects for the year for which the ABCC index was computed, fixed effects for country, log total urban population in 1800 from (25),log ruggedness from (26), and the log of the region's area.

generations of academics. On February 5th, 1835, the Lincoln Mechanics' Institute received a bust of Isaac Newton (born in 1642 in Lincolnshire) from a wealthy benefactor. To celebrate, the 19 year old son of the society's curator (and local shoemaker) gave a lecture on the "Life and Discoveries of Newton." (33). The young man, George Boole, would become the founder of modern algebraic logic.

Finally, this inspiration could be embedded in local culture. Pierre de Fermat (1605–1665), one of the greatest French Mathematicians, member of the Academy of Castres, was born in a small village, Beamont-de-Lomagne. His working life was spent in Toulouse at the Parliament (a court). Today, Beamont-de-Lomagne has a statue of him, a street named after him, a tourism office located in the house where he was born, and a yearly *fête des maths* in his honor. Every year kids learn to like mathematics at this festival.

Råshult is the name of trolley sold by IKEA, but it is also a village in Småland, Sweden, notable as the birthplace of the "father of modern taxonomy," Carl Linnaeus (1707–1778). Råshult has a monument to him, a reconstruction of the cottage where he was born, and garden based on his famous *Adonis Stenbrohultensis*, in which he first used taxonomy to classify every plant in his father's garden.

Even the medieval scholar Pierre Abelard (1079–1142), is honored in his hometown, the tiny Breton village of Le Pallet, with both a street name and a statue. His intellectual influence, philosophical writings, and his tragic romance with Héloïse (resulting in a son named Astrolabe and the castration of Abelard by an angry uncle) have left a lasting impact over several centuries.

Table 4 tests a key part of this inspiration mechanism: that growth is related to the connection between a scholar and his region of birth. The output of scholars weakly attached to their home region is not associated with growth. Because we control for more strongly attached scholars, we are indirectly controlling for any omitted variables that increased the demand for or supply of scholars. We thus argue the lack of association suggests that it is scholars influencing development in their home region, not vice versa. Moreover, the measures of weak attachment that we use are particularly relevant for our inspiration mechanism. It is hard to picture a Catholic region building monuments to a Protestant emigrant, or a scholar who died young encouraging the next generation.

Overall, we use new data and methods to show that the production of knowledge by scholars mattered for economic growth. It did so across all fields of academic research, although certain fields such as science had an outsized impact. Moreover, it had a local impact on development through the connections between a scholar and their place of birth. We hypothesise that this is because scholars can serve as an inspiration for future generations, encouraging the pursuit and application of knowledge.

Materials and Methods

For each scholar with a WorldCat Identities page, we collected the tag cloud of their "Associated Subjects" (excluding the persons who are honorary members). We then drop subjects associated with fewer than 30 scholars or that are about a specific country (e.g. "French history"). This leaves us with 1,360 subjects and 16,149 scholars with at least one subject.

WorldCat gives each subject a font size based on the relative importance of the term. We quantify the importance of a subject from 1–5 based on the rank of its font size. Thus for each scholar i and subject j, we have weights $\gamma_{ij} \in \{0, 1, 2, 3, 4, 5\}$. We then construct a data matrix Γ of dimensions $1,360 \times 16,149$ containing every γ_{ij}. Each row is an academic, each column a subject.

The k-means algorithm treats each row of Γ as the coordinates point in a $1,360$-dimensional space. It partitions the data into k clusters, minimizing the total within-cluster sum of squared deviations (TWCSS). This is the sum of squared deviation of each point from the centroid of its cluster.

k-means must be estimated using numerical methods as there is no closed-form solution. We use the default R package which implements the Hartigan-Wong algorithm (34). This algorithm starts with random guesses for the centroids of each cluster and then iteratively improves the centroids until a certain convergence threshold is reached. As the improvements converge to a local optimum, not a global optimum, we repeat the estimation 500 times, picking the replication with the lowest TWCSS.

The choice of k can be made using various criteria. We minimize the Bayesian information criterion (BIC): $TWSS_k + \log(I)Jk$, where $I = 16,149$ and $J = 1,360$. This is minimized at $k = 10$. More details in Appendix I.

Sadly, the 2 millions pages of the WorldCat Identities project were suddenly retired in March, 2023. This is bad news for those interested in measuring human capital from publications data. For the future however, we found a viable alternative using statistics drawn from the VIAF platform. See (35) for more details.

The online appendix explains how readers will be able to access the data used in the paper.

ACKNOWLEDGMENTS. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program under grant agreement No 883033 "Did elite human capital trigger the rise of the West? Insights from a new database of European scholars." We thank Julius Koschnick and our colleague at IRES for comments on an earlier draft.

References

1. Mokyr J (2002) *The Gifts of Athena: Historical Origins of the Knowledge Economy.* (Princeton University Press, Princeton).
2. Mokyr J (2016) *A Culture of Growth.* (Princeton University Press, Princeton).
3. Galor O (2022) *The journey of humanity: The origins of wealth and inequality.* (Penguin).
4. Wootton D (2015) *The invention of science: A new history of the scientific revolution.* (Penguin UK).
5. Mitterauer M (2010) *Why Europe?: the medieval origins of its special path.* (University of Chicago Press).
6. Henrich J (2020) *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous.* (Penguin UK).
7. Weber M (1930) *The Protestant ethic and the spirit of capitalism.* (George Allen & Unwin Ltd, London).
8. Cantoni D, Yuchtman N (2014) Medieval universities, legal institutions, and the commercial revolution. *The Quarterly Journal of Economics* 129(2):823–887.
9. Hill AW (1915) The history and functions of botanic gardens. *Annals of the Missouri Botanical Garden* 2(1/2):185–240.
10. De Courson B, Thouzeau V, Baumard N (2023) Quantifying the Scientific Revolution. *Evolutionary Human Sciences* pp. 1–37.
11. Johnson ND, Thomas AJ, Taylor AN (2023) The Impact of the Black Death on the Adoption of the Printing Press.
12. Grajzl P, Murrell P (2023) A Macroscope of English Print Culture, 1530-1700, Applied to the Coevolution of Ideas on Religion, Science, and Institutions.
13. Almelhem A, Iyigun M, Kennedy A, Rubin J (2023) Enlightenment Ideals and Belief in Science in the Run-up to the Industrial Revolution: A Textual Analysis.
14. Koschnick J (2023) Breaking tradition: Teacher-student effects at english universities during the scientific revolution. London School of Economics.
15. Rosés JR, Wolf N (2021) Regional growth and inequality in the long-run: Europe, 1900–2015. *Oxford Review of Economic Policy* 37(1):17–48.
16. Cinnirella F, Streb J (2017) The role of human capital and innovation in economic development: evidence from post-malthusian prussia. *Journal of Economic Growth* 22:193–227.
17. Pedersen O (1992) *Lovers of Learning: A History of the Royal Danish Academy of Sciences and Letters, 1742-1992.* (Munksgaard).
18. McClellan JE (1985) *Science reorganized: Scientific societies in the eighteenth century.* (Columbia University Press).
19. Frijhoff W (1996) Patterns in *A History of the University in Europe. Vol. II: Universities in Early Modern Europe (1500-1800)*, ed. Ridder-Symoens Hd. (Cambridge University Press).
20. British Library (2021) Database of Italian academies.
21. De la Croix D, Docquier F, Fabre A, Stelter R (2020) The academic market and the rise of universities in medieval and early modern Europe (1000-1800). CEPR Discussion Paper No. DP14509.
22. De la Croix D (2021) Scholars and literati in european academia before 1800. *Repertorium Eruditorum Totius Europae* 5:35–41.
23. Chaney E (2020) Modern library holdings and historic city growth. University of Oxford.
24. USTC (2019) Universal short title catalogue. University of St. Andrews, https://www.ustc.ac.uk.
25. Buringh E (2021) The Population of European Cities from 700 to 2000: Social and Economic History. *Research Data Journal for the Humanities and Social Sciences* 6(1):1–18.
26. Nunn N, Puga D (2012) Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics* 94(1):20–36.
27. Porta RL, de Silanes FL, Shleifer A (2008) The economic consequences of legal origins. *Journal of Economic Literature* 46(2):285–332.
28. Mokyr J (2017) *A Culture of Growth: The Origins of the Modern Economy.* (Princeton University Press).
29. Baten J, Hippe R (2018) Geography, land inequality and regional numeracy in Europe in historical perspective. *Journal of Economic Growth* 23(1):79–109.
30. Squicciarini MP, Voigtländer N (2015) Human capital and

industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics* 130(4):1825–1883.
31. Maloney WF, Valencia Caicedo F (2022) Engineering growth. *Journal of the European Economic Association* 20(4):1554–1594.
32. Cantoni D, Dittmar J, Yuchtman N (2018) Religious competition and reallocation: The political economy of secularization in the protestant reformation. *The Quarterly Journal of Economics* 133(4):2037–2096.
33. Burris S (2022) George boole in *The Stanford Encyclopedia of Philosophy*, ed. Zalta EN. (Stanford University).
34. Hartigan JA, Wong MA (1979) A k-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28(1):100–108.
35. Curtis M, de la Croix D (2023) Measuring human capital: from WorldCat identities to VIAF. *Repertorium Eruditorum Totius Europae* 10:17–22.