Researchers and the Wealth of Nations

Cabello, Matias and Rojas, Carolina

Facultad de Ingeniería, Universidad Diego Portales, Santiago, Chile.

19 May 2016

Researchers and the Wealth of Nations

Matías Cabello* & Carolina Rojas†

This draft: June 7, 2016

Abstract

Despite the repeated claim by eminent students of economic growth that scientists and inventors have contributed to economic development, no study has yet quantified this effect using the rich historical record of great minds. Introducing a novel database of per capita researchers, we show that the history of research activity (corrected for geographical biases) predicts economic growth over the long run better than any other established growth predictor, and that this predictive power, while subject to swings, has been consistently increasing through time over the long run. These conclusions are drawn after presenting a number of facts suggesting that forces exogenous to income and population growth have determined how intensively countries have engaged in research. In contrast to a large body of literature, we find that property rights and schooling have been of minor importance for research and for economic growth through modern history. Our estimated dynamic impact of researcher densities on economic growth are very consistent through a variety of samples and regressions, based either on cross-sectional or on time-series variance. Permanently doubling the number of researchers per capita had barely an impact in 1800, but today its impact might be an increase of annualized economic growth rates of almost 1% in a 20-years span.

Keywords: Economic growth, long run, science, research, education, institutions.

JEL: N10, O11, O10, O30, O43, O47.

Contents

*Facultad de Ingeniería, Universidad Diego Portales, Santiago, Chile.
Email: matias.cabello@mail.udp.cl

†Facultad de Ingeniería, Universidad Diego Portales, Santiago, Chile.
Email: carolina.rojasga@mail.udp.cl

1 Introduction

Many students of economic growth have put research as the primary cause of economic development over the long run.[1] However, the empirical literature of long-term economic growth has not yet made use of the vast biographical data on "great minds" of science and technology to explore the degree to which it can help us understand two of the most important questions in economics: the beginning of modern economic growth and why some countries are much richer than others.

The aim of this paper is to fill this gap and investigate how the emergence of researchers through history in distinct places is associated with subsequent economic outcomes.[2] To do so, we constructed a comprehensive database of birthdates and

[1] According to Simon Kuznets, "since the second half of the nineteenth century, the major source of economic growth in the developed countries has been science-based technology." (Kuznets, 1966, p. 10). Joel Mokyr has famously elaborated on this idea. For him, it was during the Age of Enlightenment that the seeds of modern economic growth were laid by the conviction among the intellectual elite that applied science was conductive to prosperity (Mokyr, 2002). Angus Maddison wrote once that, among all "societal, intellectual, and institutional changes that had taken place over the preceding four centuries", the "most fundamental" precondition for modern economic growth was "the recognition of the human capacity to transform the forces of nature through rational investigation and experiment" (Maddison, 1991, p. 52). Dominant formal growth theory, inspired by Romer (1990), has an R&D sector acting as the engine of economic growth. See also Jones (2005) on the role of knowledge creation for economic growth.

[2] There have been various attempts to quantify scientific and intellectual activity across time (e.g. Gascoigne, 1984, 1992; Mokyr, 2005; Murray, 2003). Yet, surprisingly, the effect of intellectual activity on GDP based on the record of great minds has been, to the best of our knowledge, only studied by Jakob Madsen and coauthors in a series of recent works (Madsen, 2013; Madsen and Murtin, 2015;

birthplaces of noticeable researchers out of DBpedia, a repository containing metadata from Wikipedia, which we show to be consistent with authoritative sources of the history of science in its time-series dimension as well as in its cross-country dimension, yet preferable to other sources due to its scale, homogeneity, continuity, and international coverage. We took advantage of the continuity of the data to present day in order to estimate and apply a correction against underrepresentation of non-Westerners and non-English-speakers, avoiding—or at least substantially diminishing—the common bias that affects cross country biographical data.

Our analysis of the data confirms a number of claims of the economic history literature and provides a quantitative perspective of them. Additionally, it unveils a number of hitherto obscure, yet striking facts on research and its relationship with economic growth. We hope that these facts should substantially enrich the debate on the deep causes behind long-term economic growth and international income inequality.

The first big question to be addressed in this paper relates with causality. It is well known that wealthier countries tend to be those which spend more on research in our days, measured as expenditures over GDP or as researchers per capita. Does the causality just run from income per capita to researchers per capita, as happens with most variables that correlate with income? Or does the causality run in the opposite direction as well?

A few simple, yet powerful facts illuminate this question. One is that a large number of countries have not experienced changes in the number of researchers per capita since modern economic growth started: Most European nations and ex-colonies experienced their major changes of researchers per capita between 1500 and 1820, i.e. *before* modern economic growth began; Japan underwent its great increase of researcher as a direct result of the Meiji Restoration, thus also prior to its modern growth regime; and the selected group of other successful countries that climbed through the international research ranking once modern economic growth was already underway—South Korea and Norway, for example—are exceptions to the fact that the number of researchers per capita was established for most countries before the, so called, "great divergence" of income across the world, despite a dramatic increase in standards of living.

Later changes, which are comparatively much minor than those observed within 1500-1820, are neither correlated with economic growth. For instance, it is well known that the US, France, and Germany have grown at very similar rates during the last 100 years. Yet, the number of researchers per capita has grown in the US, has fallen in Germany, and stayed constant in France.

Hence, research has not been driven by income. On the contrary, we found that income has indeed responded to research.

An example of these dynamics is provided by the European industrialization experience. Focusing on Europe has the advantage that the data on researchers is much richer and more reliable than that of other parts of the world. Additionally,

Madsen and Yan, 2013) which focus either on Britain alone or on a small subset of countries during the pre-modern-growth period.

dealing with Europe alone is a way of overcoming any sort of bias that may result from the underrepresentation of non-Western researchers. That said, what makes Europe particularly interesting is that its geographical distribution of researchers changed dramatically between 1500 and 1820 along episodes such as the Reformation and the Counter-Reformation, which abruptly shifted intellectual activity out of southern Europe and strengthened research in other parts of Europe, especially in Britain. These episodes lead to a distribution of researchers per capita by 1820 (by 1700 it was already partially established) that did not only determine what would be the distribution of researchers in the centuries to come, but what would be the future distribution of income as well. In other words, a variety of events not particularly related with income or population—perhaps related to religion above all—gave shape to a distribution of researchers in Europe by 1820, and that distribution predicts which countries are more or less wealthy today.

A similar story can be told for the worldwide sample. As in the European case, we found that there have not been major changes in the relative distribution of researchers per capita since 1820. Actually, it has been relatively stable since the Renaissance (only with the late ascent of Japan would the Western research environment be rivaled again). And the current dispersion of income across the world turns out to be strongly correlated not just with the researcher density of 1820 but with that of the 15th century as well.

Why could researchers from so long ago matter for current income levels? The geographical distribution of researcher densities has been persistent over the very long run. Therefore, 18th century researchers per capita, for instance, act as a proxy of the researcher densities during later years, and the researcher density in a moment of time is probably a proxy of the sort of human capital needed for the adoption and implementation of "complex" technologies in an economy. In this sense, the density of researchers can be thought of as causal of different economic development across the world.

In sum, when the long term is analyzed the following conclusions on causality emerge. First, what drives research to rise and fall historically is not known for sure, but income is definitively *not* a first order determinant of research. Second, the high contemporaneous correlation between income and research observed today can be seen as the result of historically and "exogenously" given researcher densities: countries with high or low numbers of researchers per capita experienced more or less modern economic growth respectively.

Ruling out causality going from income to research as a major force has tremendous implications for our interpretation of the data. One implication has to do with a fact that has received little attention so far: the number of recorded researchers worldwide has exhibited remarkably constant growth since the 18th century. On a first glance one would have interpreted this as the result of economic growth. However, in light of the evidence that points to causality running from research to income, this constancy might be behind the fact that the frontier economies have experienced almost constant income growth rates since the modern growth regime began, consistently with theoretical growth models that exhibit diminishing returns to knowledge (Jones, 2005).

Another implication of (almost purely) monodirectional causality is that it allows us to estimate the impact of research on the economy econometrically without the need of instrumentalization and all the uncertainty derived from such approach.

So, how much has research mattered for long-term economic growth? To answer this question we relied on a variety of econometric estimates that allowed us to construct a coherent interpretation.

First we regressed 2005 log income per capita in a cross section of countries with respect to the number of per capita researchers born in previous centuries (15th to 19th centuries, respectively), controlling for initial income and other covariates. Irrespective of the century of the right hand side variables, we found an income elasticity of about 33% with respect to past per capita researchers. This means that having doubled the number of researchers per capita centuries ago—no matter exactly when—is associated with per capita income today which is about one third larger. These regressions have a large R^2. Indeed, regressing log 2005 income with respect to the 19th century researcher density *alone* produces an R^2 of 0.7, suggesting that research has been a major determinant explaining today's international income inequality.

Then, to better understand the evolution over time of the relationship between international income inequality and research we run rolling cross-country growth regressions with the initial number of per capita researchers as one of the explanatory variables. Not satisfied with cross sectional estimates alone—after all, our correction of the biases could be inadequate—, we cross-checked our estimates of the impact of researcher densities on later economic growth with panel data regressions immune to (time-invariant) cross-sectional biases, obtaining results that are equivalent to our cross-sectional estimates.

We found that the closer we get to the present, the stronger the impact of historical research on economic growth. Before the modern growth regime began, it was not significantly different from zero. Yet, since the Industrial Revolution the mean of the impact has been consistently on the rise. Now, for example, a country that had doubled its number of researchers per capita two decades ago, is since then experiencing a long-term per capita income growth rate that is about 1% greater. In contrast, a country which doubled its number of per capita researchers 200 years ago would have experienced on average 0.1% higher economic growth, so that income today would be about one third larger than without having doubled the researcher density two centuries ago, just in line with our regressions based on income levels instead of growth.

And last, we run cross-country, rolling regressions with income in a variable year as dependent variable but holding fixed the density of researchers (say that of the 17th century) and controls. The result is remarkable. We found that the more time passes, the more the international distribution of income resembles that of historical researcher densities of the 19th, 18th, 17th, 16th, and even 15th century. So, for example, the researcher density of the 17th century is more closely associated with per capita GDP of 2005 than with per capita GDP of, say, 1930. In other words we could say that history matters more every year. How can this fact be explained? We interpret it to be the result of two forces previously mentioned: first, a strong persistence in the cross-sectional distribution of researchers per capita; and second, a growing impact of researchers on economic growth.

What can we say about mass schooling and Northian institutions, the two main predictors of economic growth proposed by the empirical literature during the last decades? Our estimates are robust to the inclusion of them, but their impact gets eclipsed when researcher densities enter the regressions. Literacy ratios do not predict early industrialization—a well known fact.[3] But also the impact of schooling on later development surprisingly vanishes when competing with research capabilities. Constraints to the executive may still play a role for economic growth, we found. But the Glorious Revolution is definitively not associated with a subsequent improvement of British research output. Actually, British research took advantage over continental Europe many decades before the Glorious Revolution.

Our work is related with that of students of the theory of economic growth that have put research at the center of the modern growth engine.[4] It is also related with the empirical literature on the sources of economic growth and international income inequality in general and with the economic return of research in particular.[5] Recent works with results similar to ours include Mokyr (2016), who argues that a culture which fostered research and innovation emerged in parts of Western Europe due to the lower censorship of novel thinking compared to other parts of the world, Squicciarini and Voigtländer (2014), who find that early French industrialization of cities is not much related to broad education (literacy and schooling) but significantly related to scientific elite education (proxied by subscribers to the *Encyclopédie*), Vidal-Robert (2014), who finds that in regions of Spain where the Inquisition was more active economic growth and attitudes towards innovation have been negatively affected until modern times, and Bénabou et al. (2015b) and Bénabou et al. (2015a), who find a negative relationship across the world between innovation and religiousness.

In order to facilitate the reading of this fairly large study we highlighted our assumptions, the empirical facts, and the conclusions drawn from them. The structure of our work is as follows. Section 2 presents the database and explains the adjustment we applied as correction against regional and linguistic biases affecting

[3] See Allen (2003) and Squicciarini and Voigtländer (2014).

[4] Among many others, these include Mokyr (2002), Jacob (1997), Goldstone (2009), and Landes (2003) in economic history, and Arrow (1962), Kuznets (1966), Romer (1990), Grossman and Helpman (1993), and Aghion and Howitt (1992) in growth theory. Critiques of the science-based theory of industrialization during its early phase can be found in Clark (2012), Allen (2009), and Temin (2014).

[5] A recent review of empirical literature on the economic return of research can be found in Hall et al. (2010). The empirical literature on the sources of long term growth is too vast to be fairly reviewed here. Important recent insights are that international income inequalities are very long-lasting (Comin et al., 2010) and closely associated with genetic distance of a nations' population (Spolaore and Wacziarg, 2014). Acemoglu et al. (2001) triggered a large literature devoted to studying the role of institutions for long-term development. However, this literature has been subject to much critique (Glaeser et al., 2004, Ogilvie and Carus, 2014). Jones (2015) reviews recent findings of the growth literature with emphasis on the role of Northian institutions and (static) misallocation.

the historical record of researchers. Section 3 presents our empirical findings, and section 4 concludes. Details on the database can be found in the appendix.

2 The Data

To study how creators of useful knowledge may have affected the economy we need birthyears and birthplaces of influential researchers through history. The data must be unbiased towards specific countries or regions, it has to be generated by a homogenous process through time and space, and it has to be sufficiently large to cover the entire world back to periods preceding industrialization.

It is not obvious how to construct such a database. The history of science has been written overwhelmingly by Westerners. Therefore, virtually every source will be biased towards the West. And an important part of the detailed accounting of great minds is divided in catalogues constrained to fields, eras, geographical regions, ethnic groups, or even gender. Merging Joseph Needham's volume of Chemistry and Chemical Technology in "Science and Civilisation in China" with Roshdi Rashed's chapter on historiography of "Encyclopedia of the History of Arabic Science", and keeping a homogenous quality of researcher out of such combination, is a task well beyond our capabilities.

We opted instead for a simple but fruitful approach: we downloaded all researchers appearing in DBpedia (a repository containing structured metadata gathered from the English version of Wikipedia), filtered out erroneous data, compared resulting cross country measures with official statistics and historical sources, and corrected for biases of underrepresentation. The definition of researcher we employed corresponds to a person engaged in work aimed at increasing the stock of what one might call, in terms of Kuznets (1966), "useful knowledge". More precisely, we included people engaged in all branches of theoretical and applied science, technology, mathematics, and philosophy (excluding religious studies).

While certainly imperfect in fine detail, DBpedia provides a vast amount of information—about 40.000 researchers with birthdate and birthplace, 16.000 of them born prior to the 20th century, a number larger than that of scientists in authoritative sources like Gascoigne (1984) or the Dictionary of Scientific Biography—, including researchers from almost every corner of the planet and all fields considered useful by society. Also, researchers accounted in it should have contributed to the stock of knowledge in sufficient degree as to "deserve" a Wikipedia page. This—after controlling for geographical biases—should provide a metric that is comparable across countries in a given time.

We found that an appropriate geographical bias-correction of the number of researchers per capita can be achieved with

$$\begin{aligned}\text{Adjusted log researchers p. c.}_t = &\ \text{Unadjusted log researchers p. c.}_t \\ &+ (100 - \%\ \text{English speakers}) \times 0.0125 \qquad (1)\end{aligned}$$

where the number 0.0125 corresponds to the average of columns (2) to (7) in panel C of table A2, which is discussed in the appendix. We will refer to this adjustment as the English-correction. Unfortunately, reliable statistics on the number of English speakers are available only for a small subset of countries, mainly in Europe. Therefore, we will use this correction for analyzing the European case alone.

When making comparison among countries of the entire sample we will employ an alternative, suboptimal (less precise), correction instead. It is based on the residuals of the following regression.[6]

$$\begin{aligned}\text{DBpedia log researchers p. c.}_t = \text{const.} + \text{NSF log articles p. c.}_t \\ + \text{residual}_t, \qquad t = 1986, \ldots, 2000 \qquad (2)\end{aligned}$$

These capture the average degree of over or underrepresentation in DBpedia with respect to the density of scientific articles per capita according to the National Science Foundation (NSF).The adjusted metric, to which we will refer to as NSF-correction, is thus

$$\begin{aligned}\text{Adjusted log researchers p. c.}_t = \text{Unadjusted log researchers p. c.}_t \\ - \text{average residual of (2)}\,. \qquad (3)\end{aligned}$$

An important assumption is made in both corrections. Namely, that the degree of under or overrepresentation of a country's researchers is the same that we observe in recent times, no matter if we are dealing with researchers from the 20th or from, say, the 18th century. This assumption will certainly fail to produce reasonable statistics for periods which go back too far. Societies such as the Maya or the Phoenicians let almost no trace of their scientific achievements, and that losses are certainly not accounted by the correction here proposed. But we will risk assuming that they hold as a valid approximation since the modern period or, at least, since 1820. We will hence state this assumption in two forms.

Assumption 1. *Countries with underrepresented researchers in DBpedia in recent decades also suffer from the same proportional underrepresentation through time …*

a) …since the 15th century. (Strong form of assumption 1)

b) …since 1820. (Weak form of assumption 1)

We will work with assumption 1a in order to make use of the data back to the 15th century. However, the important results should hold if 1b alone is true.

A detailed account on the construction of the dataset and on how it compares to alternative sources can be found in the appendix.

[6] The relative lack of precision has to do with a high variance of the residuals. For instance, Portugal and Spain, or Sweden and Denmark, despite having similar economies and cultures, may appear with very different correction factors. Nonetheless, the correction is important to rule out a bias against underrepresented countries or regions, as it is clearly the case for East Asia.

3 The Facts

3.1 From stagnation to steady-growth in research

Figure 1 shows the worldwide evolution of what we will henceforth refer to as the "number of researchers": the number of researchers aged 20 to 70, including deceased ones.[7] We cannot say much about the number of researchers prior to 1400. Figures are inevitably imprecise for such remote dates and world aggregates are hard to interpret in the "unglobalized" society that was in place before the great voyages. But one point seems safe to make: prior to the 15th century there were few researchers and there was almost no growth trend in the aggregate series.

This stands in sharp contrast with what we see after around 1720. The dotted line in the figure indicates a 1.58% growth trend which almost perfectly fits the series of aggregated researchers. This magnitude may not tell us too much, because it might be that the data is subject to "depreciation"—older researchers appearing less in the historical record than more contemporaneous ones, meaning that this growth rate might be an overestimation. What is clear, however, is that the number of researchers worldwide has been growing in a steady-state fashion since the onset of industrialization.[8]

Fact 1. *In contrast to pre-modern times, the number of famous researchers as a world aggregate has been increasing since the Renaissance. It has so in a steady-state-looking fashion since the early 18th century.*

What drives and what implies a constant growth rate of researchers? At a first glance, one obvious candidate explaining constant research growth is that income and (hence also) population have grown during the modern growth regime. The opposite, more provocative, interpretation is that rather income and population have grown as a consequence of research growth. The theoretical underpinning for such possibility is laid down clearly in Jones (2005), where it is proposed that the growth rate of a frictionless knowledge-absorbing economy must be proportional to the growth rate of the number of researchers in the world.[9]

Conjecture 1. *The constancy of growth of worldwide famous researchers has not been caused by constant economic growth. The causality goes in the opposite direction.*

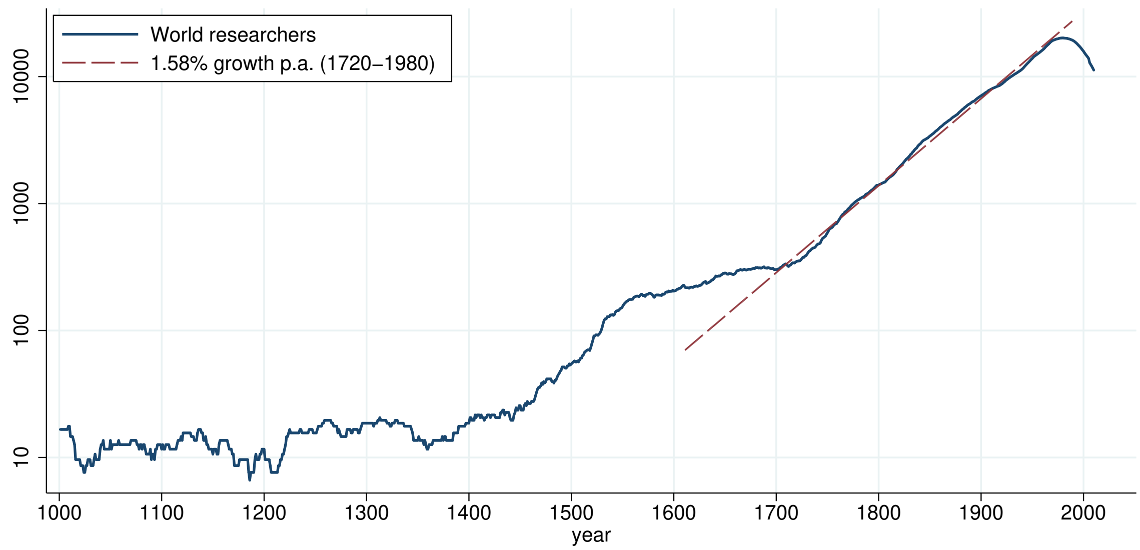
[7] Correction (3) is applied in this case. The uncorrected data has an almost indistinguishable shape. We chose 20 and 70 as upper and lower bounds for two reasons. First, almost all famous researchers have produced their contributions within these ages (see e.g. Jones et al., 2014). Second, while most researchers were productive in their thirties (we could have chosen 30 to 40 as cutting points, for instance), using a narrow age interval would lead to a high variance in the time series for early dates or countries with few researchers. In contrast, an index of researchers aged 20 to 70 provides a smooth metric which can be more sensibly employed in econometric analysis.

[8] This pattern is found in various other databases of scientists as well (see Gascoigne (1992) and our data appendix).

[9] In such case one may conjecture that the constancy of growth among industrialized economies (which is about 2% p. a.) is due to constant growth of the total number of influential researchers, with an implied elasticity larger than unity if 1.58% is an upper bound for the actual growth rate of researchers.

Empirically, one way to show that conjecture 1 is indeed a possibility, is to find that the growth rate of researchers is not primarily determined by income and population growth.

Figure 1: Steady growth of researchers worldwide



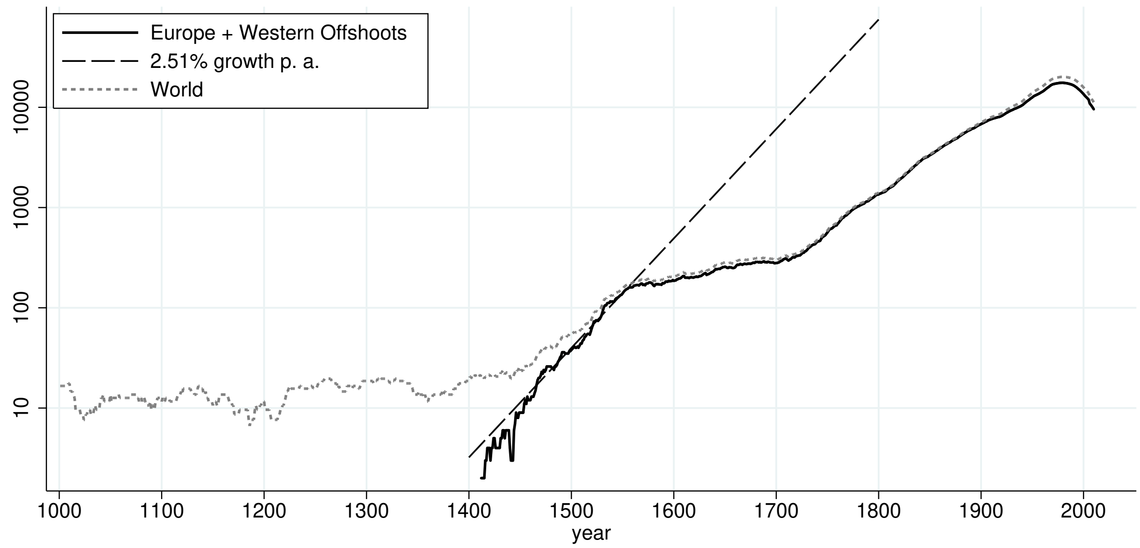
Sources: Own computations out of DBpedia.

3.2 Research and income: what causes what?

Today countries with more income per capita have more famous researchers. One interpretation for this correlation is that causality goes from higher income per capita to more research output, because higher living standards are associated with higher productivity and permit to sustain larger proportions of the population in "unproductive" activities such as research. In Malthusian societies—where productivity gains are eventually transformed into population growth—the number of researchers is expected to grow at the same rate that the overall population does: more people means higher chances of someone coming up with a good idea. And in modern economies we expect both, population and income to be causal of more researchers. Is what we see in figure 1 simply the result of these two mechanisms—income and population growth stemming out of unexplained productivity growth? The answer is no.

Figure 2 shows in log scale the number of researchers in Europe and the so-called Western Offshoots—the US, Canada, Australia and New Zealand—through time and compared to the worldwide number of researchers shown previously. Apart from the fact that this group of Westerners has dominated the world of famous researchers since the 16th century (previously it was at around 25%, in line with its

Figure 2: It has happened before!



Sources: Own computations out of DBpedia.

share of total population), the most interesting fact is that researchers of the European Renaissance grew at an even higher rate than that observed after 1750. We know well that neither population nor income grew at modern rates during the Renaissance. Hence, population and income cannot be the main determinants of research output.

Fact 2. *During the European Renaissance the number of famous Western researchers grew at rates similar to those observed during the modern growth regime.*

The notion that researchers are proportional to the overall population—an assumption usually to be found in the literature—is also refuted by the data. As shown in figure 3a, the ratio of famous researchers to overall population—the "density of researchers"—has not been constant through history. It also does not seem that income has driven the rising ratio of researchers over population, because that would imply a structural brake when modern growth began. Rather, we observe rising growth of per capita researchers since the late middle ages.

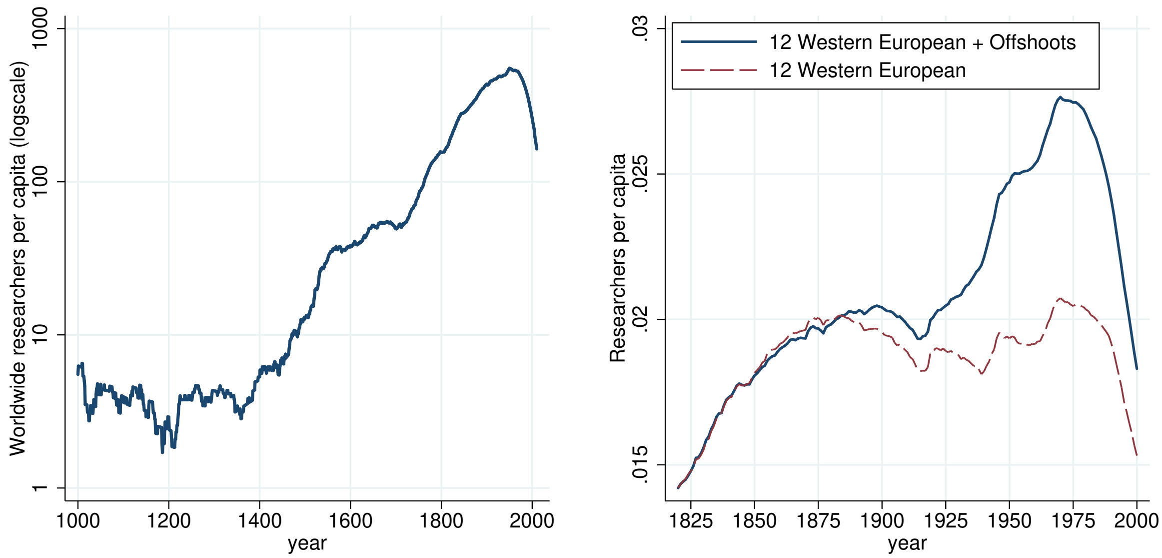
Fact 3. *Over time, the number of worldwide famous researchers per capita has neither been constant nor proportional to living standards.*

Since research growth has been primarily a Western phenomenon, figure 3b restricts the aggregation to the 12 core West European countries [10] and the Western Offshoots. One could have expected to find similar developments of research among them since they experienced similar standards of living. Yet, we can state the following.

[10] Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, United Kingdom.

Figure 3: Researchers per capita

(a) Worldwide (logscale) (b) In the core West



Notes: Per capita researchers corresponds to the number of researchers aged 20-70 over billions of world population. World population is log-linearly interpolated based on data from McEvedy et al. (1978), Biraben (1980), and United-Nations (1999).

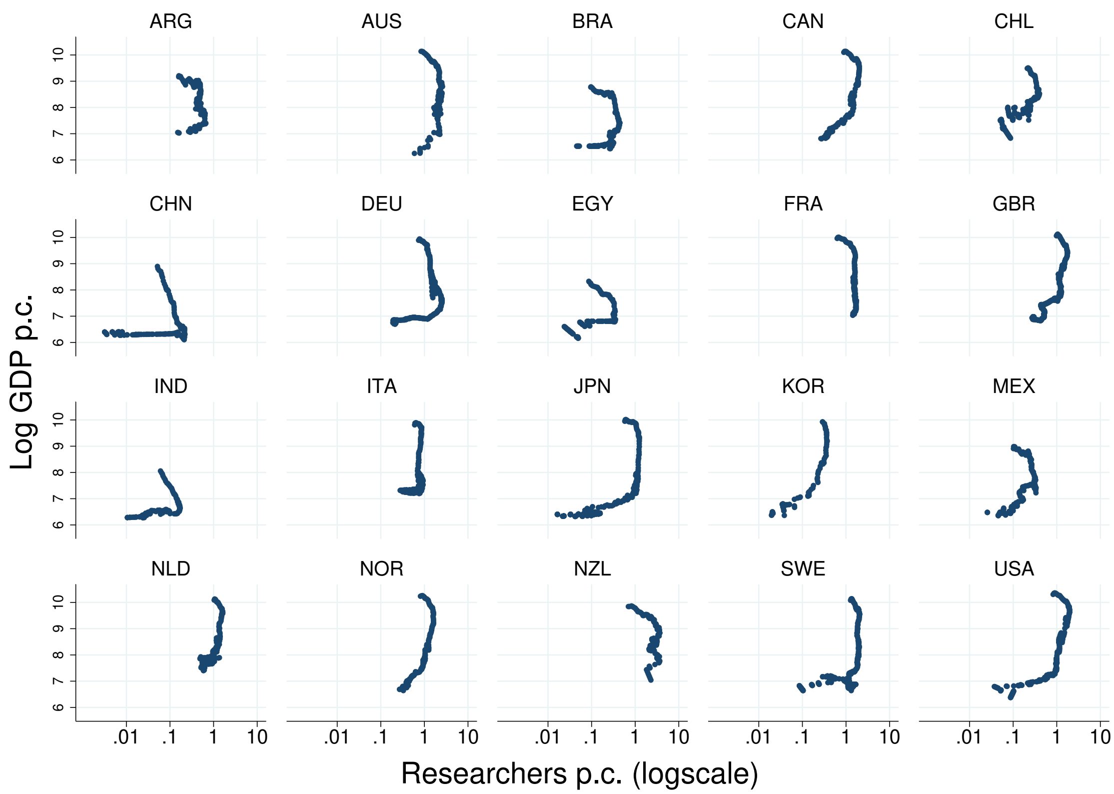
Fact 4. *Since about 1870, the density of researchers has been constant in Western Europe, while in the US it grew considerably during the 20th century .*

The development of researcher densities and income is shown at the country level in figure 4. It shows, for a number of selected economies, GDP per capita in the vertical axis and the number of famous researchers per capita in the horizontal axis, both in logs. If research would be primarily a function of income, then we would expect it to follow income dynamics. The most likely form of the figures would resemble that of Korea, with both income and research rising, or with income rising first and then research catching up. But what we observe for a large number of countries is the opposite, as for instance shown in the case of Sweden: an inverted-L form indicates that the number of per capita researchers rose at almost stable per capita income, and that income then rose at an almost stable number of researchers per capita. Most countries exhibit some variant of this inverted-L form, with researchers per capita reaching either some relatively stable number before modern economic growth started (like for France and Italy), a gentle increase (like for the US and Norway) or even a decrease (like China, India, and Germany) during the modern growth phase. It is research that changed before income, not the other way around.

Fact 5. *The density of researchers in a country was largely determined before the modern growth regime began*

Facts 2, 3, 4, and 5 let us conclude the following.

Figure 4: Per capita income and researchers since 1400



Notes: Per capita researchers corresponds to the number of researchers aged 20-70 divided by population from Maddison (2010). Population was interpolated for countries with have observations between 1500 and 1820. Per capita GDP estimates are from Maddison (2010).

Conclusion 1. *The number of researchers in a country has been exogenous to income and population, in the sense that these variables have not been first order determinants of it.*

What, then, has determined the distribution of research around the world? A look at what happened in the onset of research growth in Europe should help us to find an answer.

3.3 The renovation of Europe's research landscape

In figure 2 we saw that after 1400 something special happened to the worldwide sum of researchers, which from then on was primarily driven by developments in Europe and its offshoots. And in figure 4 we saw that the per capita number of researchers was determined mostly before modern growth began at around 1820. So what can we learn from this crucial period which precedes the modern growth regime?

Figure 5 presents in log scale the number of researchers aged 20-70 between 1300 and 1930 in a group of major European nations, corrected by the current share of

English speakers as discussed previously. Figure 6 presents the same data in per capita levels.

At a first glance, one can observe that these countries experienced notable heterogeneity in their evolution of research activity across time. But if one focuses on the developments prior to 1570, it seems that all nations where taking part of a pan-European Renaissance, with rising numbers of researchers. This was not population driven, as evidence in figure 6.

This situation came to a sudden end in the second half of the 16th century: France, the Netherlands, and Spain experienced abrupt declines in the number of researchers and witnessed substantial differences in their future developments. Italy and Germany entered periods of stagnation that lasted centuries. On the other hand, the United Kingdom was the *only* European power which maintained an almost uninterrupted rising trend from 1500 to 1700.

Why this intellectual divergence happened during this period is a challenging and important question to which we might not risk give a conclusive explanation. However, given the synchronism of slowdown and decline of researchers in continental Europe during the late 16th century, it is tempting to link this phenomenon with the religious and political turmoil of the time. As argued by Mokyr (1990, p. 76):

> The Reformation, and its natural sequel the Counter-Reformation, made Europe a more bigoted place than it had been since the Crusades: Giordano Bruno was burned by the Catholic Inquisition, Miguel Servetus by its Calvinist counterpart in Geneva. Throughout Europe in the sixteenth- and early seventeenth centuries, the authorities' patience for people who thought for themselves and were critical of dogma was wearing thin. [...] In southern Europe, which came increasingly under the domination of the reactionary power of the Counter-Reformation, the climate for technological creativity changed for the worse.

Placing the Reformation, the Counter-Reformation, and the violence associated with them at the center of the intellectual atmosphere seems to be consistent with a number of facts observed in figures 5 and 6.

First, the slowdown of Germany coincides with the beginning of the Counter-Reformation in 1545. It was only after the Thirty Years War (1618-1648) that the number of researchers began a slow but progressive recovery.

Second, Italy had experienced the devastation of its north during the Italian Wars (1494 - 1559)—what seems to have had impact on its research activity—, but its number of researchers began to *decline* continuously since the second half of the 16th century, suggesting that the Roman Inquisition—founded 1543 and lasting until 1808, being particularly repressive during the second half of the 17th century (see Mokyr, 2007, footnote 39)—, might have something to do with it.

Third, the French Wars of Religion—which extended from 1562 to 1598—coincide to some extent with a sudden decline in researchers in that country.[11]

Fourth, Italy and Germany, but to some extent also the Dutch Republic and arguably even France, where politically fractionalized regions, what made the "suppression" of influential intellectuals "by the ruling orthodoxy and vested interests more difficult" (Mokyr, 2007).[12] This might explain why these countries did not experience a collapse of researchers similar to that of highly centralized Spain, where the Inquisitional tribunal had "*de facto* power on all Spanish territories" (Vidal-Robert, 2014).

Fifth, the fall of Spanish researchers seems to be consistent with the timing of inquisitional repression. Table 1 presents the number of trials during distinct periods of Spanish inquisitional activity. The data distinguishes trials against other religions (Jews, Muslims, and Lutherans) and trials against Catholics, accused of blasphemy and superstition, among other charges. This is the kind of persecution which one expects to have been more detrimental for the rise of progressive thinking in science and philosophy. While accounting just 6% until 1520, from then on trials against Catholics made most of total trials afterwards. The number of these trials per year was particularly high between 1570 and 1620, coinciding with a stark decline of Spanish researchers in figure 5.[13]

Table 1: Trials of the Spanish Inquisition

	Period				
	1478 - 1520	1520 - 1570	1570 - 1620	1621 - 1700	1701 - 1808
Total trials	5865	6502	13874	9881	3946
Against other religions	5251	2446	6055	3859	1136
Against Catholics	351	4056	7819	6022	2810
in percentage	6%	62%	56%	61%	71%
by year	8.4	82.8	159.6	76.2	26.3

Notes: Trials against Catholics include charges of bigamy, blasphemy, superstition, fornication, and acts against the Inquisition, among others. *Source*: Own computations based on Vidal-Robert (2013).

Sixth, recent econometrical work by Vidal-Robert (2014) has shown that geographical areas within Spain where the Inquisition was more active experienced less urbanization and population growth afterwards, and that these regions currently have more conservative opinions towards scientific advances. Also Bénabou et al. (2015b,a) have recently found a negative relationship between religiosity and

[11] To have an idea of the scale of this event, the total deaths during the French Wars of Religion "has been roughly estimated at between two and four million" (Knecht, 2002, p. 91).

[12] "[F]ragmentation of power was as prevalent *within* states as between them. For one thing, power was divided between central authorities, provincial estates, and local courts. In Germany and Italy, of course, this had become formalized, but in other 'states' such as the Dutch Republic, the central government had little power." (Mokyr, 2007, p. 25, original emphasis).

[13] Between 1621 and 1700 the density of researchers remained stagnant, while the number of trials against Catholics per year was similar to that observed between 1520 and 1570, with relatively stagnant researchers as well. Then, after 1700, the number of trials declined and the density of researchers increased.

innovation (patents per capita) across states of the US and across countries, suggesting that religion might has played a role in determining attitudes towards research (and eventually income levels) systematically through history.

It should be noted that revisionist work on the Roman and Spanish Inquisitions has argued that the role played by the Inquisition shaping the differential development of science in Europe has been overstated [Kamen (2014), Tarrant (2014)]. But was it? In light of the evidence presented so far a deeper reassessment of this issue seems necessary.

And seventh, Britain, which was almost untouched by the religious turmoil affecting continental Europe, did indeed not experience any comparable reduction in its number of researchers.

Summing up we might propose the following interpretation.

Conjecture 2. *Intellectual intolerance during the Reformation and Counter-Reformation had long-lasting negative effects on the density of famous researchers in some European countries.*

How accurate this interpretation really is for interpreting the evolution of research densities needs to be proven. It is hard to believe that the intellectual history of Europe between the 17th and the 18th can be explained by a simple story. The case of the Netherlands is particularly puzzling. Why did research output decrease that much relative to other parts of northern Europe? Some historians have already pointed to this question. "It is striking", Mokyr (2000) wrote, "that in the great advances in chemistry and physics in the late eighteenth and early nineteenth century, Dutch names play less of a role than one would have expected by virtue of the country's numbers of literate, urban people." According to Jacob (1997), the reason behind this phenomenon is to be found in the lack of *interest* that both, the Dutch private sector and the Dutch government, had in the development of science. But why did the Dutch had little interest in science?

Be it as it may, we can surely state the following.

Fact 6. *In Europe, the cross-country relative distribution of researchers per capita experienced drastic changes between 1550 and 1820. Some regions of Northern Europe, especially Britain, emerged to dominate the research landscape during this period meanwhile Southern Europe fell behind.*

Figure 5: Log researchers in selected countries

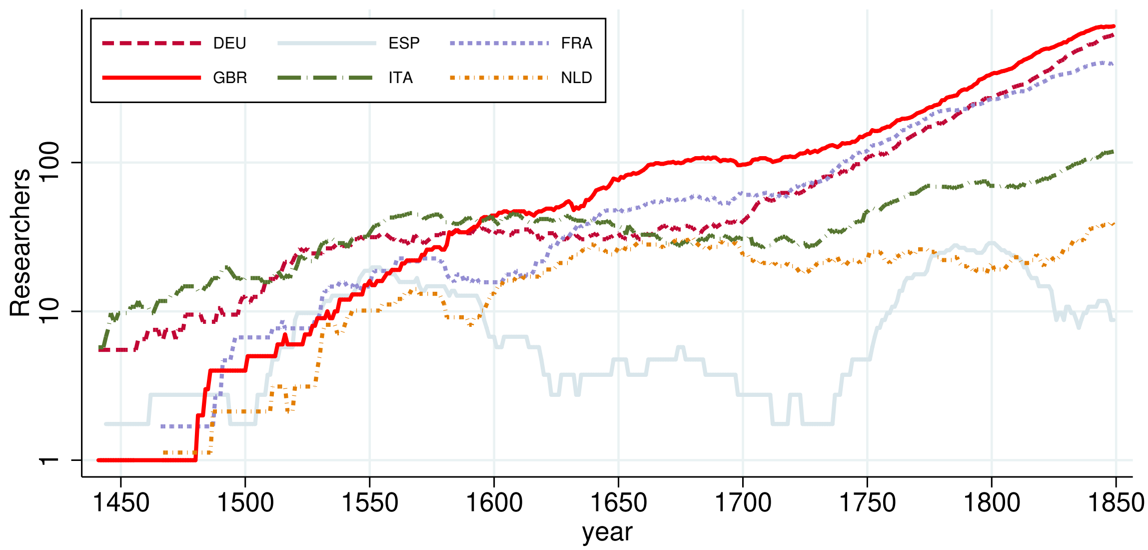
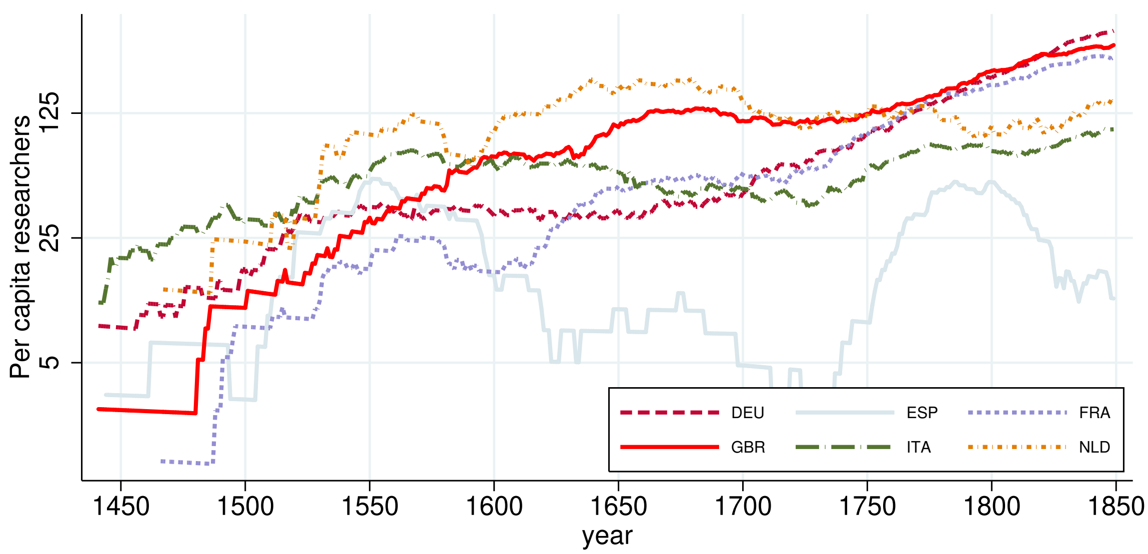


Figure 6: Log researchers per capita



Notes: The number of researchers in each country is adjusted according to the English-correction as discussed in section 2.

3.4 European industrialization

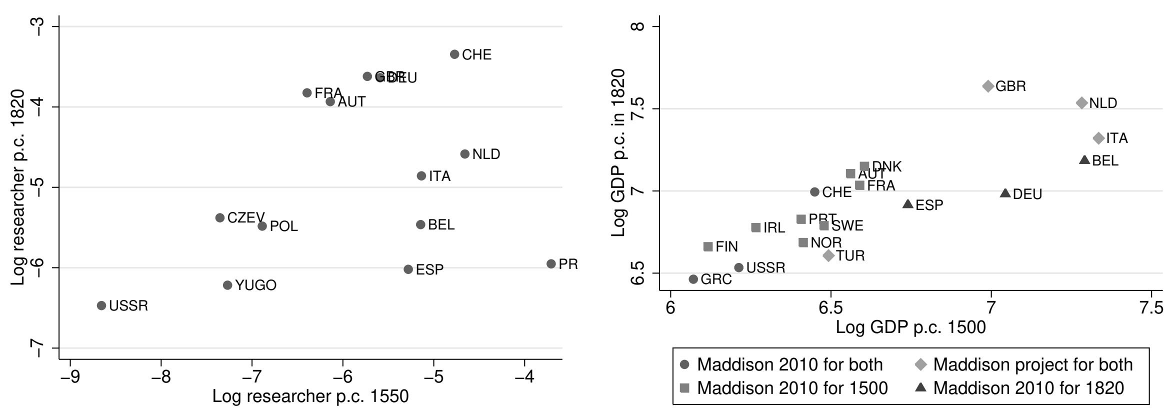
Between 1550 and the beginning of modern economic growth, which we might date to 1820, the research landscape in Europe was largely transformed. How does this resulting distribution of researchers relate to the process of industrialization that unfolded afterwards? And—before answering that—can the direction of causality between research and economic growth be identified?

To answer both questions we must start by noticing that while the distribution of researchers across Europe changed dramatically between 1550 and 1820, as shown in figure 7a, the relative distribution of living standards did not. This is shown in figure 7b, where different markers denote whether the data is from either the original work of Maddison (2010) or from revisions of 2013, so that the reader can infer which differences may be due to the underlying data construction methodology. We see that the British Industrial Revolution was already underway, the Dutch had lost its naval hegemony and Italy had been rampaged by civil wars and invasions. Yet, these are exceptions to the fact that, in general, the relative distribution of income across Europe remained mostly that of the *ancien régime* until 1820.

Figure 7: Changes between 1500 and 1820

(a) The research landscape changed dramatically between 1550 and 1820

(b) . . . meanwhile GDP remained relatively stable



Sources: See text.

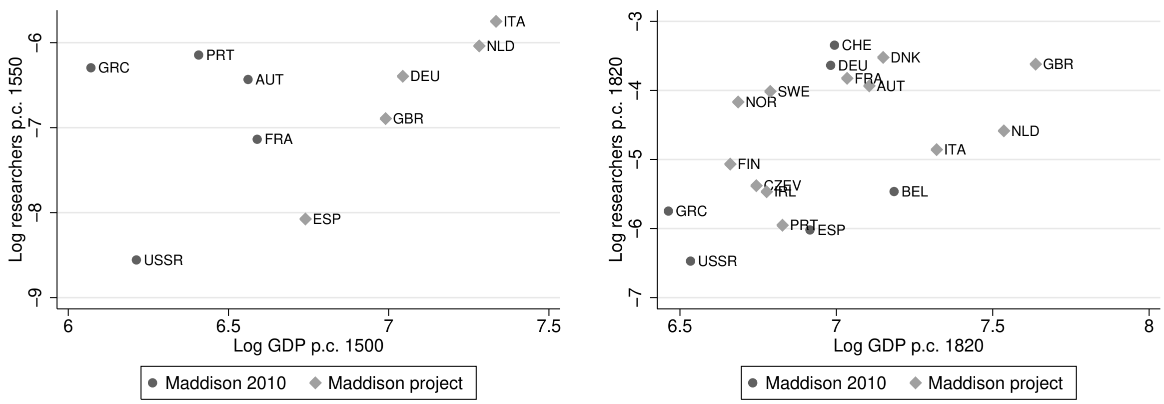
Fact 7. *Between around 1500 and 1820 the European relative distribution of researchers changed dramatically while the relative distribution of income remained relatively stable.*

Naturally, this tells us that how many researchers per capita a country had was not strongly correlated to how much income it produced. This is shown more clearly in figures 8a and 8b, where we see a positive but very disperse relationship.

Fact 8. *Prior to the modern growth regime, the European cross-country correlation between per capita researchers and per capita GDP was weak.*

Figure 8: Researchers vs. contemporaneous income

(a) Researchers vs. GDP around 1500 (b) Researchers vs. GDP in 1820

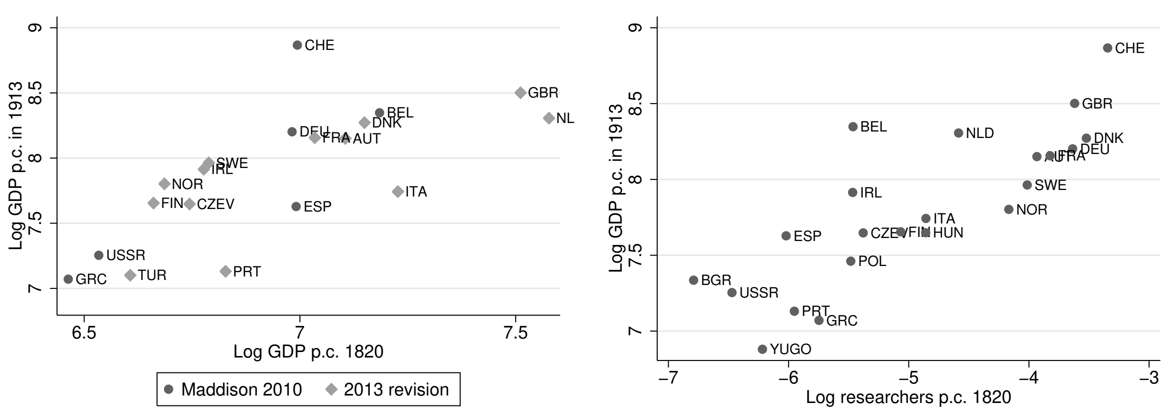


Sources: See text.

While during *ancien régime*—that long period which we might date between 1500 and 1820—relative income across European countries remained static, the scenario started to change dramatically once modern economic growth flourished around 1820. As shown in figure 9a, by 1913, after just one century of modern growth, formerly backward Switzerland, as well as Germany, France, Austria, and Denmark rapidly started to approach British standards of living. Meanwhile, Southern Europe fell behind. Indeed, in figure 9b we see that—except for Belgium—the resulting income distribution of 1913 closely mimics that of researchers back in 1820.

Figure 9: GDP and research distribution in Europe since modern growth began

(a) Relative income changed between 1820-1913 (b) . . . reflecting the researcher density of 1820



Sources: See text.

The understand the argument consider the results of table 2. Panel A of the table presents the results of the following cross sectional OLS regressions:

$$\frac{\ln(\text{pcGDP 1913}) - \ln(\text{pcGDP } x)}{1913 - x} = \text{const.} + \ln(\text{researchers p.c. } x)\beta_1 + \ln(\text{pcGDP } x)\beta_2 + (\text{Maddison 2010 dummy } x)\beta_3 + \text{residual},$$

quantifying how economic growth from 1500, 1600, 1700, and 1820 to 1913, respectively, is predicted by the initial researcher densities after controlling for initial income and a dummy indicating the source of GDP.

In column (1) we see that the predictive power of the researcher density of 1500 is statistically indistinct from zero. Yet, in columns (2) and (3) we find that the researcher densities of 1600 and 1700 have some predictive power of economic growth up to 1913. And in column (4) we see that researchers per capita very strongly predict economic development up to 1913. How do we interpret these results?

We may first turn to Panel B of the table, which presents the OLS slope coefficient of

$$\ln(\text{researchers p.c. 1913}) = \text{const.} + \ln(\text{researchers p.c. } x)\beta + \text{residual}.$$

In column (1) we see that the relationship is very weak between researcher densities of 1500 and 1913, meaning little persistence of relative research activity across Europe. It is stronger and more precise in columns (2) and (3), taking 1600 and 1700 as staring dates respectively. And it is very strong in column (4), meaning that between 1820 and 1913 the relative distribution of researchers per capita across Europe remained rather stable.

This stability suggests a first interpretation of the results of panel A. Not old researchers but contemporaneous researchers do facilitate economic growth. Yet, the stability of researcher densities through time makes it appear as if older researchers matter.

Alternatively, older researchers may do matter, in the sense that they permeated society with "cultural" traits conductive to growth, perhaps through the education system or through a high valuation of researchers by the society. Be it as it may:

Fact 9. *In the cross section of European countries, the number of researchers per capita in 1820 is positively associated with early industralization.*

But why should European countries with a more research-oriented culture had grown more than others? After all, knowledge is (at least partly-) non-rival. Hence, if we want to explain economic development, who (which country) invents or discovers new productive knowledge does not matter; it rather matters who (which country) is able to adopt the new ideas.

We believe, though, that the ratio of researchers over population acts as a proxy of both, technological creation capabilities and technological *adoption* capabilities of a society. The first point is obvious. The second has to do with the kind of society which gives rise to a high number of researchers. We expect such society—or at

Table 2: Growth is predicted by research in Europe

A) Dep. var.: anualized p.c. GDP growth for the period indicated (× 100)

	(1) 1500-1913	(2) 1600-1913	(3) 1700-1913	(4) 1820-1913
Initial log p.c. researchers	-0.00940 (0.0353)	0.133*** (0.0330)	0.167** (0.0548)	0.337**** (0.0759)
Initial log p.c. GDP	0.0174 (0.209)	-0.525*** (0.139)	-0.713** (0.290)	-0.328 (0.196)
Maddison 2010 dummy	0.0242 (0.149)	0.0493 (0.0750)	0.0556 (0.103)	0.262 (0.166)
Observations	10	13	12	18
R^2	0.014	0.650	0.740	0.633

B) Dep. var.: researchers per capita in 1913

	(1) 1500-1913	(2) 1600-1913	(3) 1700-1913	(4) 1820-1913
Initial log p.c. researchers	0.256 (0.219)	0.354 (0.205)	0.361** (0.123)	0.604**** (0.0909)
Observations	12	17	15	22
R^2	0.077	0.183	0.363	0.721

Notes: OLS estimates with robust standard errors in parenthesis. Significance at the 10%, 5%, 1%, and 0.1% levels are indicated by *, **, ***, and **** respectively. *Source*: See text.

least the intellectual elite of it—to have a comparatively high level of human capital and the capacity to understand novel ideas from around the world.[14]

That said, the exact mechanism through which a higher researcher density might lead to higher economic growth is not fully understood by us yet. Therefore may propose this interpretation as an assumption for the remainder of this paper.

Assumption 2. *The number of famous researchers per capita in a country or region serves as a proxy of the intellectual capabilities of a society to adopt new technologies.*

Summing up, the relative distribution of researchers across Europe was mostly determined prior to 1820. After the modern growth regime took off, countries with higher researcher densities experienced higher economic growth rates and the ranking of income levels finally resembles that of per capita researchers.

What we have shown here is similar in spirit to a Granger causality test. The research environment changed in a certain way, which was independent of income, and income then reflected those changes. We may conclude as follows.

[14] See also Griffith et al. (2004).

Conclusion 2. *Across Europe, a higher relative number of researchers per capita, if maintained over time, caused higher economic growth in a country once the modern growth regime began in 1820.*

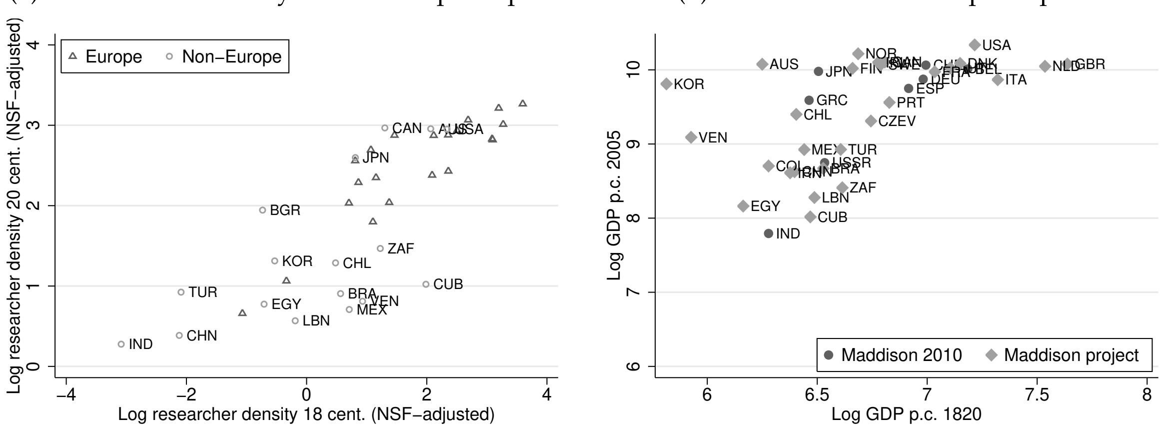
3.5 A world shaped by persistent researchers

Conclusion 2 has important implications for what we expect to observe in other samples. If the geographical distribution of researcher densities has remained relatively constant since 1820, then we should be able to predict international modern economic growth rates with the 1820 researcher densities. If the geographical distribution of researcher densities has been relatively constant since even earlier dates, then the densities of these earlier dates should be able to predict modern growth as well.

We can see in figure 10a that for both, the European and the non-European samples, research allocation has been similarly persistent from the 18th to the 20th century. And in figure 10b we see that this persistence is greater than that of income.

Figure 10: Worldwide GDP and research distribution since modern growth began

(a) 20th vs. 18th century researchers per capita (b) 2005 vs. 1820 income per capita



Sources: See text.

Despite similarities since 1820, the European and worldwide stories will differ substantially if we go further back to the 15th century. Panel A) of table 3 shows the estimated correlation based on the following cross-country regression:

$$\log \text{20th c. born researchers p.c.} = \text{const.} + \beta \log \text{i-th c. born researchers p.c.}, \quad i = 19, 18, 15$$

Columns (1) to (3) are restricted to the European sample. They show us that in the continent there has been persistence of the density of researchers since the 19th and 18th centuries, but not since the 15th century. This stands in contrast with columns (4) to (6), where persistence of researcher densities across the world sample goes back even to the 15th century. This means that, while within Europe events such as the Reformation and Counter-Reformation had enough strength to largely overwrite the previous order, such changes must look minuscule in terms of the global distribution of researchers.

Limitations of the number of recorded researchers born prior to the 15th century do not allow us to go further back in time at a country level. It is clear, though, that prior to the Renaissance the persistence should have been much weaker. Most noticeable researchers during the Middle Ages were Muslims, and during the antiquity they were Chinese, Greeks, and Indians. But since around 1500, today's world order of research—with Europeans and their descendants dominating the research distribution—was already established.[15]

To be careful, it must be noted that relative living standards and technological sophistication have been shown to be remarkably persistent as well.[16] Hence, anything that shows persistence through time—researcher densities included—may does so as a result of any persistent unknown force that is driving income and everything that may come with it. To be sure that research is special indeed, and not just as persistent as income, panel B) of table 3 shows the correlation of per capita GDP in 2005 with per capita GDP of 1913, 1820, and 1500. What we see is that for the world sample the persistence of researchers is greater than that of income.[17] The same holds for Europe in column (1) and (2). Only in column (3) we observe more persistence of income than of researchers. But we saw previously that this has to do with the great changes of research that took place in the continent between the 15th and the 19th century, and that these changes actually gave shape to the future income distribution that would emerge in Europe between 1820 and 1913, and which roughly lasts until our days.

Fact 10. *The relative distribution of researchers for the worldwide sample has been strongly persistent since the 15th century. Its persistence is higher than that of per capita income.*

It must be no surprise, then, that in figure 11 log income levels in 2005 across the world strongly correlate with the log density of researchers going back to the 19th, 18th, 17th, 16th, and even 15th century.

Fact 11. *A strong correlation is observed between current international income levels and the historical densities of famous researchers as old as from the 15th century.*

Putting these facts together we may conclude as follows.

[15] In words of Needham (as quoted in Maddison, 1997): "When we say that modern science developed only in Western Europe in the time of Galileo during the Renaissance and the scientific revolution, we mean, I think, that it was there alone that there developed the fundamental bases of *modern* science, such as the application of mathematical hypotheses to Nature, and the full understanding of the experimental method, the distinction between primary and secondary qualities and the systematic accumulation of openly published scientific data".

[16] See Comin et al. (2010), Spolaore and Wacziarg (2013), and Nunn (2014).

[17] If income is measured with substantially more error than researcher densities, then this conclusion does not hold. But if researchers are measured with substantially more error than income—the most likely case, we believe—then the fact that researchers show more persistence that income is even more telling.

Table 3: The geographical persistence of research output

A) Per capita researchers born in the 20th century vs. those born in previous centuries

	European countries			All countries		
	(1) Born 19c.	(2) Born 18c.	(3) Born 15c.	(4) Born 19c.	(5) Born 18c.	(6) Born 15c.
Correlation	0.916****	0.775****	0.179	0.930****	0.838****	0.659***
	[23]	[19]	[13]	[54]	[36]	[15]

B) Per capita GDP in 2005 vs. per capita GDP in previous years

	European countries			All countries		
	(1) 1913	(2) 1820	(3) 1500	(4) 1913	(5) 1820	(6) 1500
Correlation	0.760****	0.528**	0.529*	0.779****	0.575****	0.503*
	[23]	[19]	[13]	[54]	[36]	[15]

Notes: Observations in brackets. Significance at the 10%, 5%, 1%, and 0.1% levels are indicated by *, **, ***, and **** respectively.

Conclusion 3. *Modern economic growth across the world can be predicted with the historical density of researchers, because its distribution has remained relatively stable since then.*

Two notes of care may deserve mention with respect to fact 11 and conclusion 3. First, we must remember that the quantitative relationship obtained is dependent on assumptions 1a and 1b, where the strong form 1a is much less likely to hold. Second, due to constraints on researcher and population data we left out some parts of the world which, given the small sample of countries with researchers born in the 15h century, could easily leverage the estimated slope.

That said, it must be emphasized that conclusion 3 conditions the prediction on persistence. If a country had a large researcher density during the pre-modern-growth period but a low researcher density afterwards, we expect it to exhibit low growth during the modern growth regime. Conversely, if a country had a low researcher density prior to 1820 but a large researcher density afterwards, we expect it to have experienced high economic growth rates during the modern growth era.

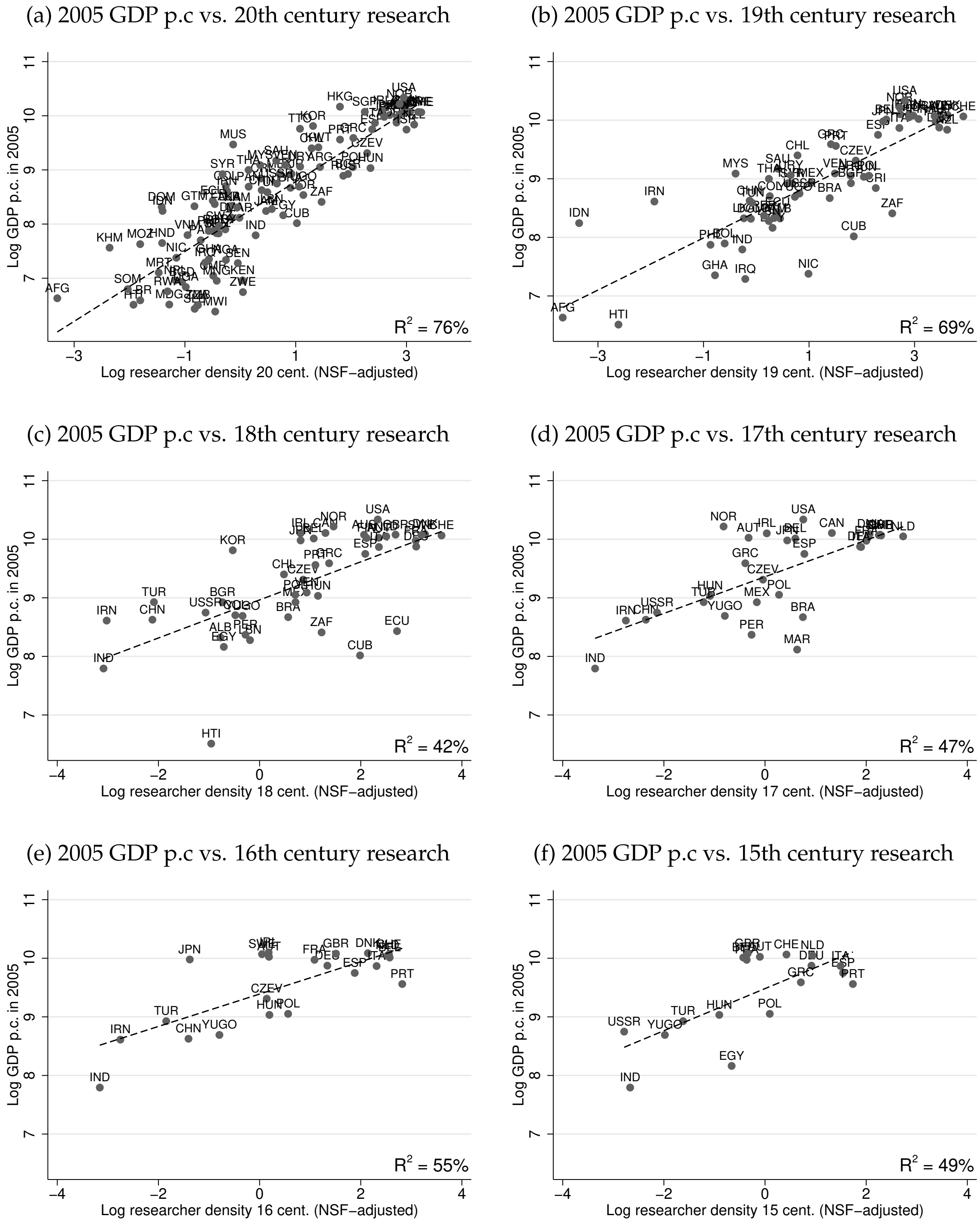
One further conclusion might be drawn from figure 11. Just how important has the "culture of research" been for international economic growth? As presented in figure 11b, the regression of 2005 log GDP per capita with respect to log per capita researchers born in the 19th century has an R^2 of about 70%. If it is taken into account that most researcher densities are historically determined largely by historical accidents independently of income, then the fit suggest that about 70% of current international income inequality may be due to unequal research intensities—or, more precisely, due to what researchers per capita are actually proxying for.[18] Admittedly,

[18]Of course, this metric is not filtering out the distortion brought about by measurement error nor

this would be a definitive conclusion only if the explanatory variable is not just exogenous, but irrefutably a first order determinant of income, and uncorrelated with other relevant factors as well.

that the research environment may have changed in the meantime, as the points of Indonesia (IDN), Cuba (CUB), and South Africa (ZAF) illustrate. In that sense, 70% might be interpreted as a lower bound if no "third factor" is causing a spurious correlation.

Figure 11: 15th to 20th century researchers vs. 2005 income levels



Notes: Country labels are standard ISO 3166-1 codes, except for the aggregates of the former Soviet Union (USSR), Czechoslovakia (CZEV) and Yugoslavia (YUGO). The straight line corresponds to the OLS fit. *Sources*: See text.

3.6 Estimates of the researchers' impact

Figure 11 already hints that historical densities of researchers must be related to the uneven economic development across the world. What can we expect from a policy promoting research in our days?

A first approximation the answer is given in our baseline cross sectional estimates shown in table 4. Each column contains the OLS estimate of the cross-sectional relationship between 2005 GDP per capita as dependent variable and the historical densities of famous researchers born in previous centuries, after controlling for a constant, initial income, a dummy indicating if initial income is from Maddison (2010) instead of the 2013 revision, a dummy controlling for the spectacular rise of East Asia, and one for the collapse of Eastern Europe.[19] Formally:

$$\ln \frac{\text{GDP}_{2005}}{\text{population}_{2005}} = \text{const.} + \beta \ln \frac{\text{researchers born}_{i\text{th-century}}}{\text{population}_{i\text{th-century end}}} + \text{controls} + \text{residual} \qquad (4)$$
$$i = 15, \ldots, 19$$

One can notice that the coefficient of the researcher density is relatively homogenous in magnitude—it makes little difference whether we use 15th or 19th century researchers, because little changed in the relative distribution of famous researchers after 1500—and highly significant. The estimates tell us that a country with twice the number of per capita historical researchers is expected to have about 30% more income today.

It should be noticed that doubling the number of researchers is not a utopian task. Norway doubled its density first from 1860 to 1900 and then again from 1900 to 1980. And, echoing the Meiji Reformation, Japan incremented its density of famous researchers between 1880 and 1920 by an impressive factor of 10.[20] Even the US came close to doubling its density of researchers between 1930 and 1980. However, we should also emphasize that these post-1800 examples are exceptions to the rule. Most countries did not experience substantial changes in their densities after 1800. Indeed, the persistence of the densities over time explains why the coefficients of table 4 are so close to each other. Many countries which climbed to top positions in research densities even 5 centuries ago still remain at the top and have higher per capita incomes than those that lagged behind in terms of research densities.

Fact 12. *In the cross section of countries, the elasticity of per capita income today with respect centuries-old per capita researchers is about 0.3.*

But when, exactly, did the action happen? To explore the time-variant role played by researchers for economic growth, we run the following rolling regressions, where

[19] Controlling for initial income is essential since it is well known that income has been remarkably persistent through centuries (Comin et al. (2010)). We considered pertinent the inclusion of a dummy for East Asia because of the remarkable achievement of the region, which may be a result of other factors (notwithstanding the important role of research). We also decided to control for Eastern Europe because of the still lasting economic impact of the Soviet collapse.

[20] These figures are in terms of the (NSF-articles) adjusted number of per capita researchers. Similar results are obtained with unadjusted numbers.

Table 4: Baseline cross-sectional regressions

	Dep. var.: per capita income in 2005				
	15c.	16c.	17c.	18c.	19c.
Log p.c. researchers	0.307*	0.377***	0.260**	0.250****	0.310****
	(0.157)	(0.110)	(0.0929)	(0.0646)	(0.0618)
Log p.c. GDP	1.256*	-1.201**	0.637	0.554*	0.454****
	(0.566)	(0.541)	(0.540)	(0.275)	(0.125)
Maddison 2010 dummy	0.611	-0.275	0.255	-0.00447	-0.612*
	(0.455)	(0.255)	(0.194)	(0.129)	(0.305)
East Asia dummy	0	0.457	0.0923	0.763**	0.729****
	(.)	(0.459)	(0.392)	(0.315)	(0.141)
Eastern Europe dummy	0.322	0	-0.0659	-0.0283	-0.117
	(0.566)	(.)	(0.259)	(0.104)	(0.102)
Observations	15	17	25	36	54
R^2	0.620	0.646	0.534	0.594	0.784

Notes: OLS estimates with robust standard errors in parenthesis. Significance at the 10%, 5%, 1%, and 0.1% levels are indicated by *, **, ***, and **** respectively. *Sources*: Per capita GDP is latest available figure from Maddison (2010) or, if available, from the 2013 updates of the Maddison Project (Bolt and Zanden, 2014) or Broadberry (2013).

t denotes the end year of economic growth and d the number of years of the period analyzed:

$$\left(\frac{\frac{\text{GDP}_t}{\text{pop}_t}}{\frac{\text{GDP}_{t-d}}{\text{pop}_{t-d}}} \right)^{\frac{1}{d}} \cdot 100 - 100 = \text{constant}_t + \beta_{t,d} \ln \frac{\text{researchers}_{t-d}}{\text{pop}_{t-d}} + \text{controls} + \text{residual} \quad (5)$$

$$t = 1600, \ldots, 2010$$
$$d = 20, 50, 100.$$

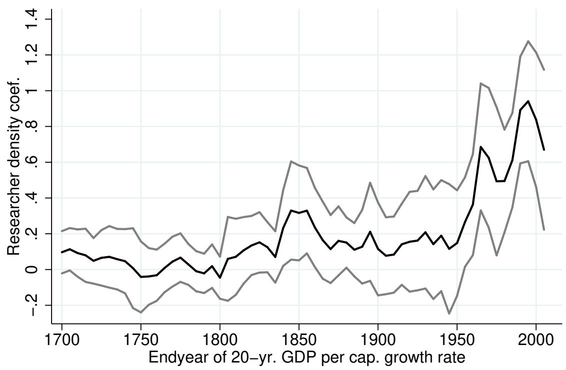
Since our focus is on the long-run component of GDP variation, we used HP-filtered ($\lambda = 1000$) annual data available in the Maddison Project (Bolt and Zanden, 2014) or Broadberry (2013).[21] Also, in order to include a large number of observations prior to 1820, we log-linearly interpolated values before applying the HP-filter. Figure 15 shows the interpolated and filtered data for a selected group of countries with particularly long time series.

As controls we implemented the same as in table 4. The resulting $\beta_{t,d}$ are plotted in figure 12. They tell us that the predictive power of researcher densities has been unstable over time, with swings that can last over decades. They also tell us that, since about 1800, the predictive power of researcher densities has been consistently increasing over the long run.

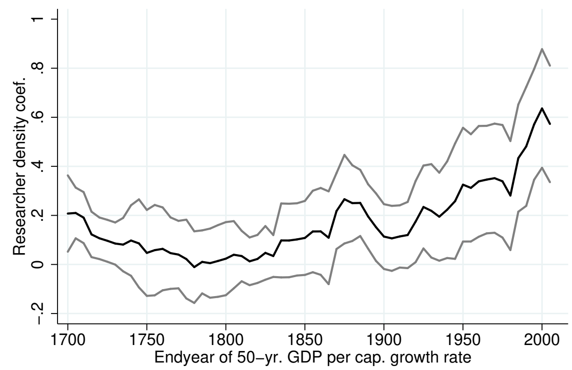
[21] In addition we included GDP data from Maddison (2010) for France, Switzerland, Denmark, Austria, Norway, and the Easter European 7 aggregate.

Figure 12: Impact of log research density on long-term income growth rates

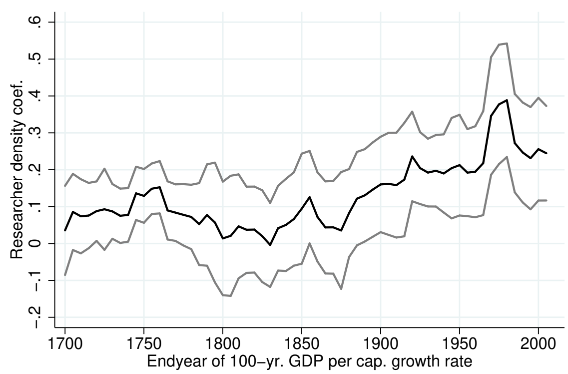
(a) $d = 20$



(b) $d = 50$



(c) $d = 100$



Notes: The plotted coefficient is the one of log research intensity of equation 5 with 95% confidence intervals. Population is from Maddison (2010) and per capita GDP is from Bolt and Zanden (2014). Missing values were replaced by the log-linear interpolation.

One way of partially avoiding the reliance on our cross-sectional correction against overrepresentation is to exploit the time-series dimension of the data. This is done in our second set of baseline estimates, presented in table 5. The functional form of the OLS estimates is the following:

$$\left(\frac{\frac{\text{GDP}_{i,t}}{\text{pop.}_{i,t}}}{\frac{\text{GDP}_{i,t-d}}{\text{pop.}_{i,t-d}}}\right)^{\frac{1}{d}} \cdot 100 - 100 = \text{const.} + \alpha_t + \alpha_i + \beta\left(\ln\frac{\text{researchers}_{i,t-d}}{\text{pop.}_{i,t-d}} - \ln\frac{\text{researchers}^{\star}_{i,t-d}}{\text{pop.}^{\star}_{i,t-d}}\right)$$
$$+ \tau t\left(\ln\frac{\text{researchers}_{i,t-d}}{\text{pop.}_{i,t-d}} - \ln\frac{\text{researchers}^{\star}_{i,t-d}}{\text{pop.}^{\star}_{i,t-d}}\right)$$
$$+ \text{other controls} + \text{residual} \qquad (6)$$
$$t = 1700, \ldots 2010 \qquad d = 20, 50, 80, 100$$

Here i denotes the country, t denotes the year and d denotes the length of the growth period considered (20, 50, 80, and 100 years). The dependent variable is again based on the interpolated and HP-filtered data mentioned previously, and is presented for a number of countries and $d = 50$ in figure 16.[22]

In addition to country fixed effects (α_i) and time fixed effects (α_t), we controlled for initial income as has become usual in the empirics of growth. We also controlled for lagged growth, because long term growth is indeed persistent and a spurious correlation could arise from its omission.

The term $\frac{\text{researchers}^{\star}_{i,t-d}}{\text{pop.}^{\star}_{i,t-d}}$ denotes a benchmark similar to a research-density-frontier, and corresponds to the average density of famous researcher among Italy, the UK, France, Germany, and the Netherlands.[23] Figure 17 shows the research gap for a selected group of countries with particularly long time series. The coefficient of interest are β, which captures the effect of filling the research gap, the time interaction τ, and the coefficient of lagged income growth.

The resulting long term multipliers for the years 1820, 1913, and 2010 are shown at the bottom of the table. The they tell us that the estimated impact went from close to zero in 1820 to values that range from about 0.3 to about 0.7 in 2010, depending on the value of d.

We suspect that the different impact one finds using longer or shorter intervals of time simply reflects the fact that other growth-relevant events happened with greater probability in larger time-intervals, eclipsing the role of past research.

A more flexible functional form for the time-varying impact is shown in figure 13, now based on a 10th degree time-polynomial. Again, we see that the impact of research densities on income growth increase trough time, corroborating our cross-sectional results.[24]

In sum, the panel estimates are very consistent with those obtained with the cross sectional rolling estimates presented in figure 12.

Fact 13. *Based either on cross sectional or on within country panel estimates, the elasticity of income growth with respect to the relative number of researchers has been increasing since 1820, when it was close to zero. For 20-years growth it has risen to about 0.7% in our days, while for 100-years growth it has risen to about 0.3%.*

What could explain an increasing importance of researchers for subsequent economic growth? For Maddison (1997), it was the "gradual infiltration of the scientific

[22]The figure shows that growth rates over 50 years have not been constant. Instead, they have been gradually increasing during the modern growth regime, in contrast to Kaldor's famous observation of balanced growth. This is one reason for including time fixed effects. Another reason is that time fixed effects might also account for research depreciation: older researchers can be systematically underrepresented in the dataset.

[23]It is not necessarily the frontier, since some nations have surpassed this benchmark at times, but it has lower variance than the actual frontier and it is, we believe, a relevant yardstick to compare the research capabilities of any country. Population to construct the researcher densities is taken from Maddison (2010). We used log-linear interpolations to fill missing population values.

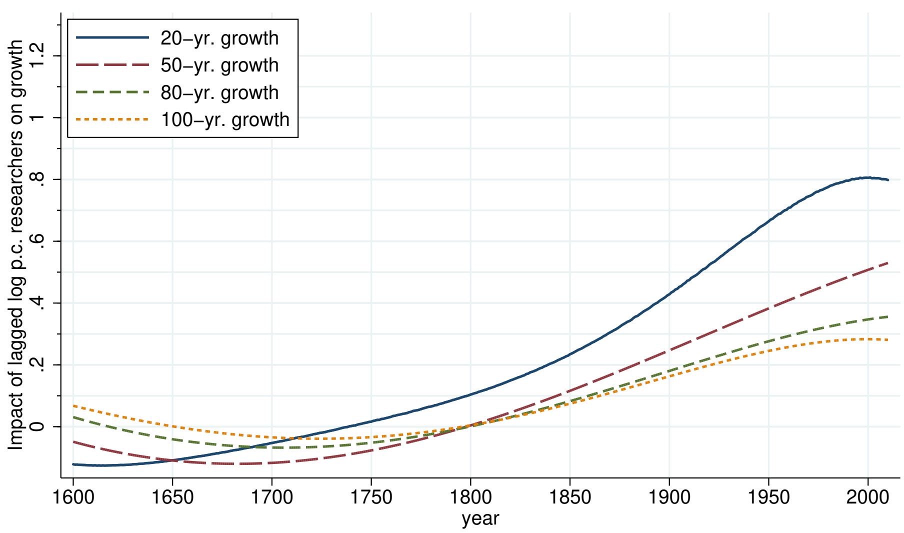
[24]The only important additional insight is that the researchers' impact on 20-years economic growth has been weakened since about the 1970s—an interesting fact indeed, but which we will not explore in detail now.

Table 5: Baseline panel estimates

Dep. var.: Annualized growth rate of HP-filtered per capita GDP ($\times$100)	(1) 20 yr.	(2) 50 yr.	(3) 80 yr.	(4) 100 yr.
laggedLogResGap	-6.221****	-4.355****	-3.007****	-2.398****
	(0.822)	(0.130)	(0.0732)	(0.0565)
laggedLogGDP	-2.033****	-1.590****	-1.188****	-1.047****
	(0.312)	(0.0507)	(0.0257)	(0.0226)
interaction	0.00344****	0.00236****	0.00165****	0.00133****
	(0.000433)	(0.0000686)	(0.0000399)	(0.0000306)
Impact in 1820	0.04	-0.05	-0.00	0.02
Impact in 1913	0.36	0.17	0.15	0.15
Impact in 2010	0.70	0.40	0.31	0.28
R^2	0.59	0.83	0.90	0.94
Observations	11277	8734	7035	6185

Notes: The regressions include country-fixed-effects and time-fixed-effects. Missing GDP and population observations are filled with log-linear interpolation. Data for the UK and the Netherlands was inferred from the contribution of Great Britain and England, and of Holland respectively. The regressions start with $t = 1700$. We omitted two countries, Iraq and Libya, because of their unusual behavior of GDP, which was in both cases clearly hit by exogenous shocks. *Sources*: See table 4.

Figure 13: 10th-order time polynomial

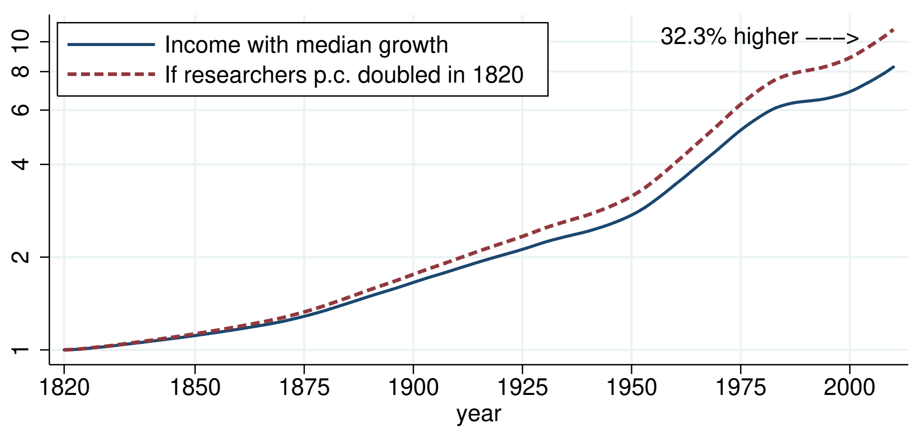


Notes: See text.

approach into educational systems". For Mokyr (2009) and Goldstone (2009) it was the increasing interaction between science and production from the 18th century onwards. Also, we would like to add, it could have been the result of knowledge accumulation which, after reaching a breaking point, spilled over to production. Before 1800, scientific knowledge available to mankind was relatively poor compared with later standards and certainly did not provide much guidance to technological improvements.[25] But technology eventually became increasingly dependent on scientific advances. This gradual change, we believe, may have been merely the result of scientific knowledge being "big enough" as to spill over to economically useful applications. Once technologies started to depend on science, the number of people in a country capable of understanding the principles that govern technology should have started to matter for technological adoption. And a higher density of researchers should be a proxy of the density of people with such capacities.

How does this dynamic effect relate to our baseline cross-sectional estimates of the elasticity of today's income with respect to past researcher densities presented in table 4? A simulation of the implied effect is resented in figure 14. We use the 100-years-growth estimate, because it arguably captures better the long-term effect of the researcher density. The simulation shows how a country which would growth at the median growth rate per year could have benefited from doubling the number of per capita researchers in 1820. Its current income level would in such case be about 30% larger than without the increased researcher densities, just in line with our baseline cross-sectional estimates of the elasticity of current income levels with respect to past researcher densities.

Figure 14: Simulation out from column (4) in table 5



Notes: The benchmark income level is computed as 1 times the median annual growth factor of the HP-filtered interpolated GDP series.

Fact 14. *Panel estimates controlling for past income and country and time fixed effect also imply an elasticity of current income with respect to past researcher densities of about 30%.*

[25]See Mokyr (1990).

Figure 15: Interpolated per capita GDP

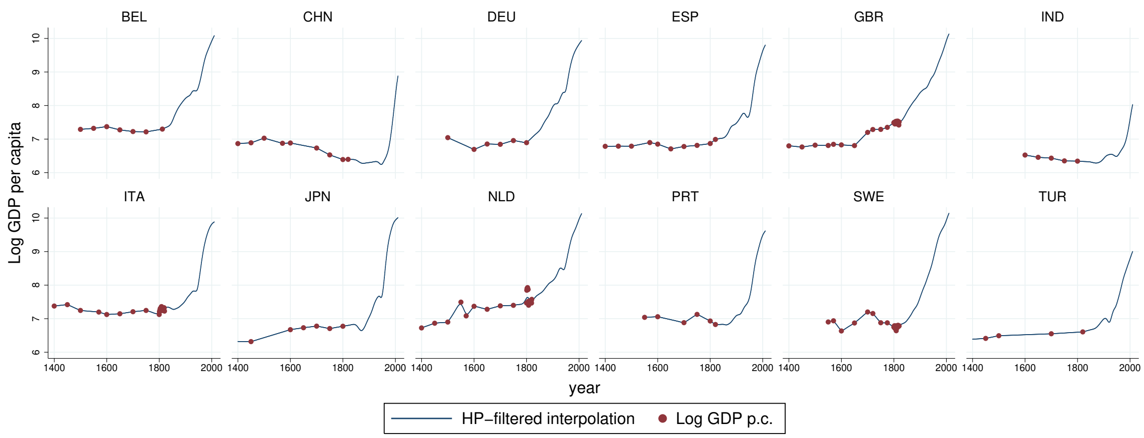


Figure 16: 50 years per capita income growth

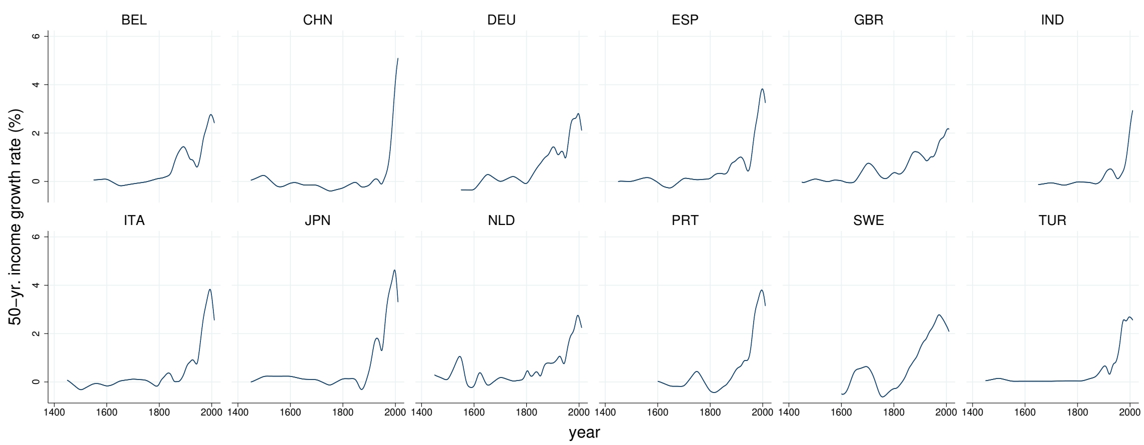
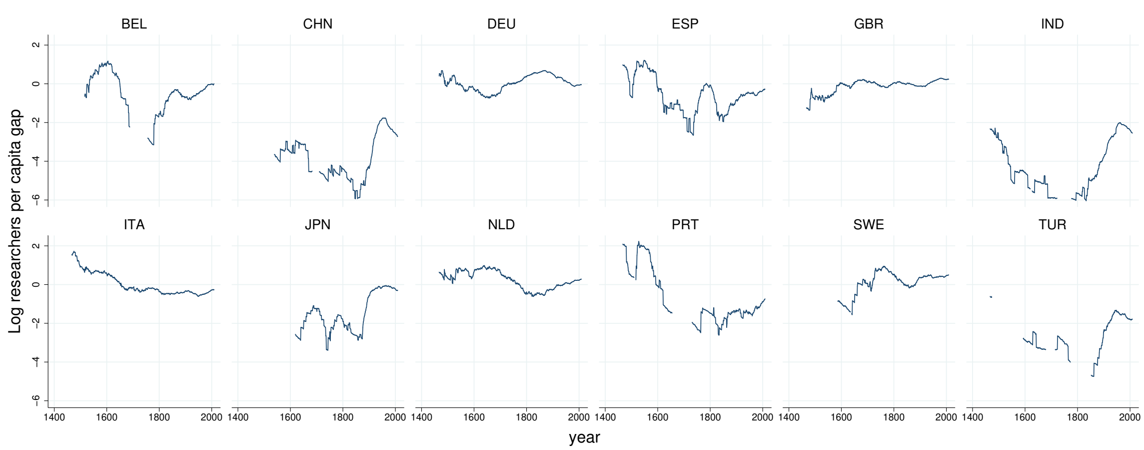


Figure 17: Research gap



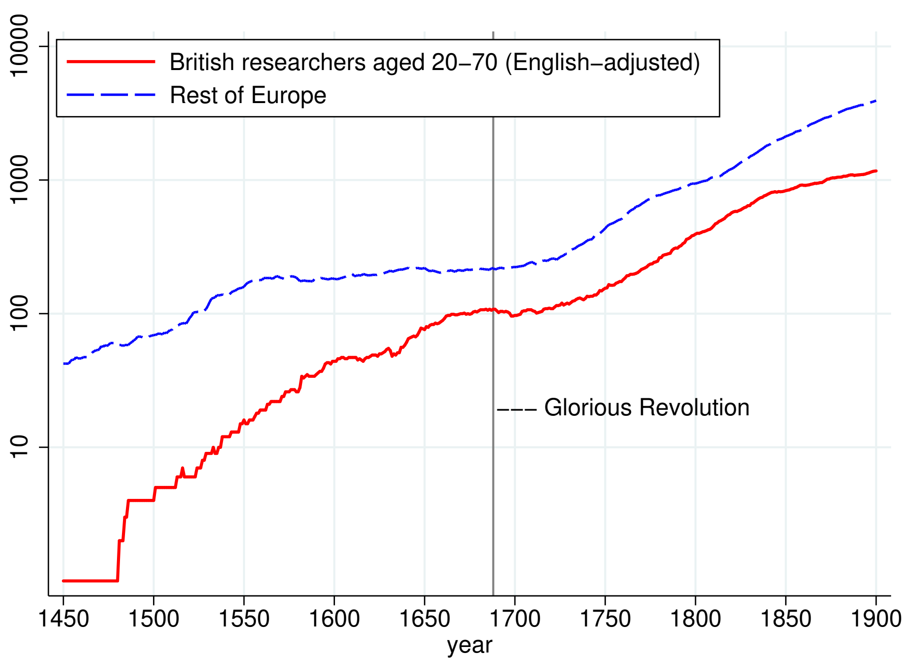
3.7 On Northian institutions

In recent decades, the institutional view of economic growth advanced by Nobel Laureate Douglass North and popularized by Acemoglu and Robinson (2013) has gained great attendance in the scholarly debate. Given its prominence, we studied the relationship that exists between famous researchers and some Northian propositions.

The central claim of Northian institutionalism is that rich economies grew prosperous mainly due to their effective protection of property rights. Without fearing to lose their property, so the argument, people in these societies had greater incentives to engage in investment and innovation, eventually leading to prosperity.

One of intitutionalists' specific propositions is that the industrial revolution started in Britain as a consequence of its Glorious Revolution of 1688. According to Acemoglu and Robinson (2013), after the Glorious Revolution, "[t]he government adopted a set of economic institutions that provided incentives for investment, trade, and innovation. It steadfastly enforced property rights, including patents granting property rights for ideas, thereby providing a major stimulus to innovation." (p. 102)

Figure 18: The Glorious Revolution and research activity



Notes: . *Source*: Own computation based on DBpedia.

If so, one could expect to see this reflected in the number of famous researchers that appeared at that time in Britain. Figure 18 shows the evolution of the number of researchers aged 20 to 70 in the United Kingdom and in the rest of Europe. Prior to the Glorious Revolution, the number of famous researchers in the UK grew at higher rates, both compared to its later performance as well as compared to their continental counterparts. Did the Glorious Revolution have a positive impact on British research output? The answer is a clear no.

Fact 15. *Compared to its previous growth performance and that of Continental Europe, British research was not positively affected by the Glorious Revolution.*

With respect to research, at least, the Glorious Revolution was not what made Britain strong. Rather, it was the fact that, for whatever reason, British research grew when continental Europe stagnated during the 17th century.

Institutionalist have further proposed that a higher degree of constraint to the executive should lead to more economic growth, finding some empirical support (Acemoglu and Johnson, 2005, Glaeser et al., 2004). To assess whether research systematically predicts growth once accounting for the institutional setting, we run our baseline regressions, now including as a control the index of constraints to the executive, which is available starting in 1800.

In figure 19 we present the results obtained after including the executive constraint as a control to the cross-sectional rolling regression (5). The upper row plots the coefficient obtained for the researcher density, showing that the estimated trends and magnitudes of the researchers' impact are robust to the inclusion of the institutional control. Meanwhile, the constraints to the executive, plotted in the lower row, appear not significantly distinct from zero.

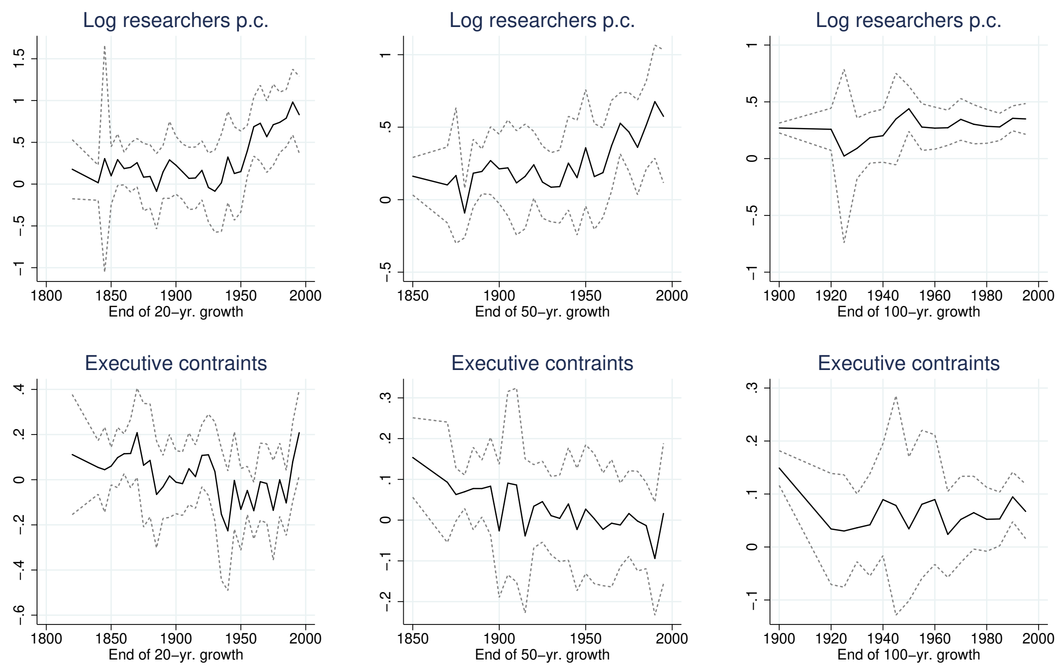
A similar exercise is presented in table 6, now with our baseline panel regression including as control the index of constraints to the executive. Again, trends and magnitudes of the researchers' impact are robust to the inclusion of the institutional control.[26]

Conclusion 4. *Researcher densities predict modern growth after controlling for constraint to the executive.*

[26] With regard to the coefficient of the institutional variable, the results are partly contradictory. We suspect that part of the significance in the panel estimates is spurious. Constraints to the executive have risen over time in most countries, and so have economic growth rates too. But perhaps both trends are driven by an exogenous factor. On the other hand, the cross sectional estimates of constraints to the executive, a categorical variable, may show up not significantly different from zero due to the imperfect measurement, what makes it hard to take a stand at all about the role that this variable may has or not for economic development.

Figure 19: Coefficients and confidence intervals of rolling cross-sectional estimates

(a) $d = 20$ (b) $d = 50$ (c) $d = 100$



Sources: The executive constraint index is from the Polity IV database, version 2013 (Marshall et al., 2013).

Table 6: Panel estimates controlling for constraints on the executive

	(1) 20 yr.	(2) 50 yr.	(3) 80 yr.	(4) 100 yr.
laggedLogResGap	-6.630****	-5.070****	-3.959****	-2.576****
	(1.617)	(0.326)	(0.322)	(0.253)
interaction	0.00365****	0.00268****	0.00210****	0.00143****
	(0.000843)	(0.000172)	(0.000168)	(0.000129)
laggedXconst	0.0170	0.0149***	0.0170****	0.0310****
	(0.0132)	(0.00488)	(0.00360)	(0.00320)
Impact in 1820	0.01	-0.18	-0.14	0.02
Impact in 1913	0.35	0.07	0.05	0.15
Impact in 2010	0.70	0.33	0.26	0.29
R^2	0.62	0.81	0.88	0.93
Observations	7246	4630	3133	2383

Notes: The regressions also control for country-fixed-effect, time-fixed-effects, and initial per capita GDP. See details in table 5. *Sources*: The executive constraint index is from the Polity IV database, version 2013 (Marshall et al., 2013).

3.8 On human capital

Human capital—habitually measured as average years of schooling or average literacy—has been widely used as an indicator of how well positioned a country is in order to absorb technical change. The number of researchers per capita in a country, we would like to argue, should as well provide an index of (upper-tail) human capital and knowledge *absorption* potential—although it is naturally a measure of knowledge *creation* as well. The idea behind this claim is that researchers must be familiarized with the last development in their field, and hence should have the capabilities to understand and possibly facilitate the adoption of new techniques developed somewhere in the world. As researchers are commonly involved in teaching as well, countries with high research density are likely to have also a high density of particularly well educated population, capable of implementing new and old techniques and even creating new ones. On the other hand, countries with high levels of education should be more likely to produce famous researchers as well.

In fact, we found that the number of researchers per capita is strongly correlated with two standard measures of human capital: schooling enrollment and years of schooling, as exemplified in figures 20a and 20b.[27]

One natural interpretation of these close relationships is that high levels of mass education are conductive to a high density of researchers. However, there might be also a reverse causality—from researchers to mass schooling—, something to which we will return below.

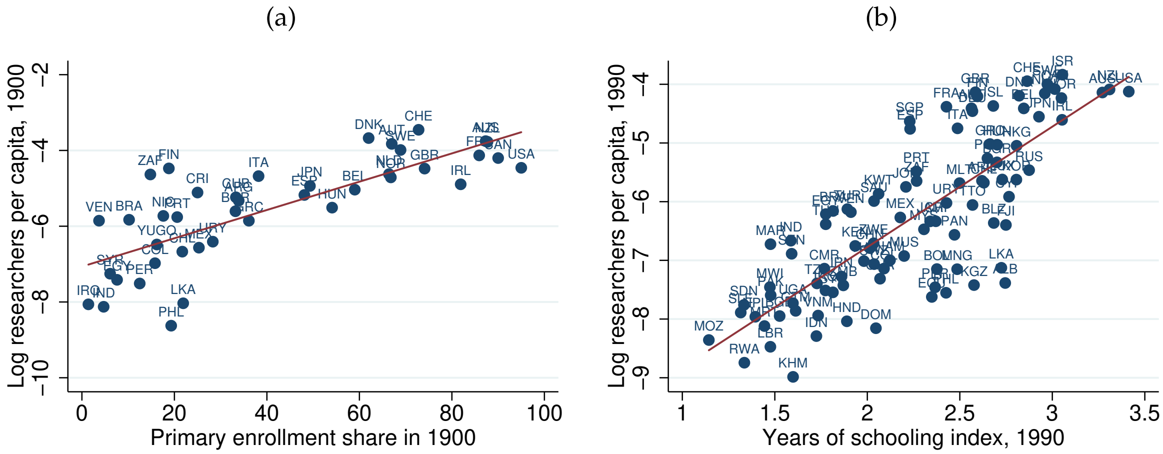
Fact 16. *Since 1900, measures of human capital are strongly correlated with densities of researchers across countries.*

Does the strong correlation between schooling and researchers we observe since 1900 also hold prior to the industrial revolution? Cross country data of literacy and years of schooling simply does not exist. What we might use instead is a raw proxy of literacy: the percentage of people able to sign their name. Figure 21 presents this measure for a small number of pre-modern economies for which this measure is available. One can see that the per capita number of researchers and people signing their name are rather uncorrelated.[28] This might be simply the result of scarce data. Alternatively, it rather might been telling us that these measures correspond to different things: the literacy proxy being a measure of broad education and the density of researchers being a measure of upper-tail human capital. The latter interpretation is consistent with recent work by Squicciarini and Voigtländer (2014), who find in their analysis of 18th century French city-level data that subscriber density to

[27]The cross country correlation between the log per capita researchers aged 20 to 70 in 1900 and schooling enrollment rates in the same year taken from Benavot and Riddle (1988) is over 0.8 (figure 20a). One also finds a similar correlation between the log of active per capita researchers and the human capital index of the Penn World tables 8.0 (figure 20b). These indicators are not just spuriously correlated through GDP: the residuals of their regressions with respect to log GDP are significantly correlated too.

[28]The correlation coefficient between the literacy indicator in 1500 and researchers per capita in 15th century is 0.23 (with a p-value of 0.56) and the one literacy indicator in 1800 and log researchers per capita in 18th century is 0.52 (with a p-value of 0.15).

Figure 20: Research intensity vs. human capital



Notes: Researchers per capita in 1990 are those aged 20 to 70 in that year. Country labels are standard ISO 3166-1 codes, the line corresponds to the OLS fit. We used primary enrollment of England and Wales as a proxy of the one for the UK. *Sources*: Primary enrollment rates of 1900 are from Benavot and Riddle (1988). Human capital indexes of 1990 are from the Penn world tables 8.0. Populations are from Maddison (2010). Researchers are from DBpedia.

the *Encyclopédie* —a proxy of upper-tail human capital—is uncorrelated with literacy rates in that period. While in contemporary times the population-wide average of human capital and the upper-tail average human capital are correlated across countries, perhaps this was not the case in the past.

Fact 17. *Previous to the industrial revolution, available measures of human capital are not correlated with densities of researchers across countries.*

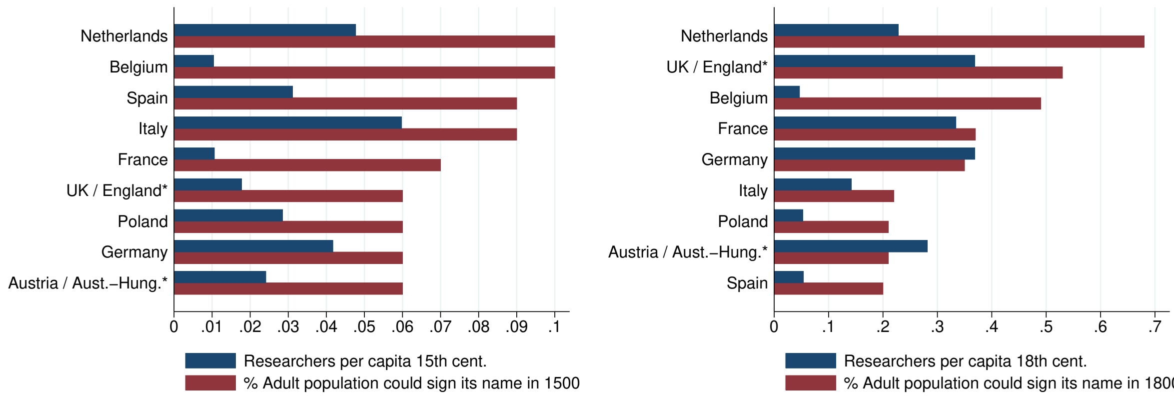
Unfortunately, we do not have sufficiently long time series of schooling to perform the rolling regression and panel exercises that were possible with the executive constraints.[29] Table 7 presents regressions similar to the one of column (5) of the cross-sectional baseline regressions of table 4, but now controlling for human capital indicators. Columns (1) and (2) show that primary enrollment in 1940 and average years of schooling in 1960—captured by the PWT8.0 index of human capital—are significant predicting 2005 income levels. However, as shown in the remaining columns, when introducing 19th century researchers both measures of average human capital turn non significant, while the impact of 19th century researchers has a significance level below 0.1% and point estimates not far from that estimated in column (5) of table 4

Fact 18. *In the cross section, the historical density of researchers remains a robust predictor of development after controlling for schooling.*

[29]Using the PWT 8.0, which covers data starting in 1950, we get results similar to those obtained in our polynomial panel regression despite issues of collinearity of the almost constant researcher densities and the country specific dummies. The results are available upon request.

Figure 21: The density of researchers and a proxy of literacy

(a) Around 1500 (b) Around 1800



Notes: * England and Austria-Hungary correspond to the geographical regions of the literacy proxy. Researchers per capita correspond to those born in the respective centuries, corrected by English speakers as discussed in section 2, divided by population at the end of the century. *Sources*: The literacy proxy is from Allen (2003). Populations are from Maddison (2010). Researchers are from DBpedia.

Combining the evidence shown so far—(i) income levels not predicting the emergence of researchers prior to the industrial revolution, (ii) no significant correlation between research density and literacy prior to the industrial revolution, (iii) high persistence of the cross country density of researchers, (iv) a strong correlation between research density and average human capital measures after the industrial revolution, and (v) research intensity strongly predicting income levels and growth rates, even controlling for average human capital.—we might adventure to put forward the following hypothesis:

Conjecture 3. *Where researchers emerged, income level rose, eventually leading to high levels of education of the society as a whole.*

If this is correct, then the causality runs, at least historically, rather from research to education than the other way around.

Table 7: Schooling vs. researchers in income regressions

Dependent variable: log GDP per capita in 2005

	(1)	(2)	(3)	(4)
Prim. enrollment 1900	0.0116***		0.00658	
	(0.00393)		(0.00510)	
Log schooling index 1960		1.137***		0.356
		(0.380)		(0.339)
Log 19th-cent.-born researchers p.c.			0.302****	0.256****
			(0.0783)	(0.0642)
Log GDP p.c. 1913	0.624****	0.650****	0.259	0.395**
	(0.162)	(0.158)	(0.173)	(0.155)
Eastern Europe dummy	-0.248	0	-0.120	0
	(0.230)	(.)	(0.209)	(.)
East Asia dummy	1.045***	0.769****	0.821****	0.571****
	(0.365)	(0.211)	(0.217)	(0.152)
Observations	48	52	43	44
R^2	0.681	0.732	0.788	0.813

Notes: Robust standard errors in parentheses. Significance at the 10%, 5%, 1%, and 0.1% levels are indicated by *, **, ***, and **** respectively. See details in table 4. *Sources*: Primary school enrollment ratios are from Benavot and Riddle (1988), the human capital index is from the PWT8.0. More details in table 4.

3.9 How GDP has taken shape

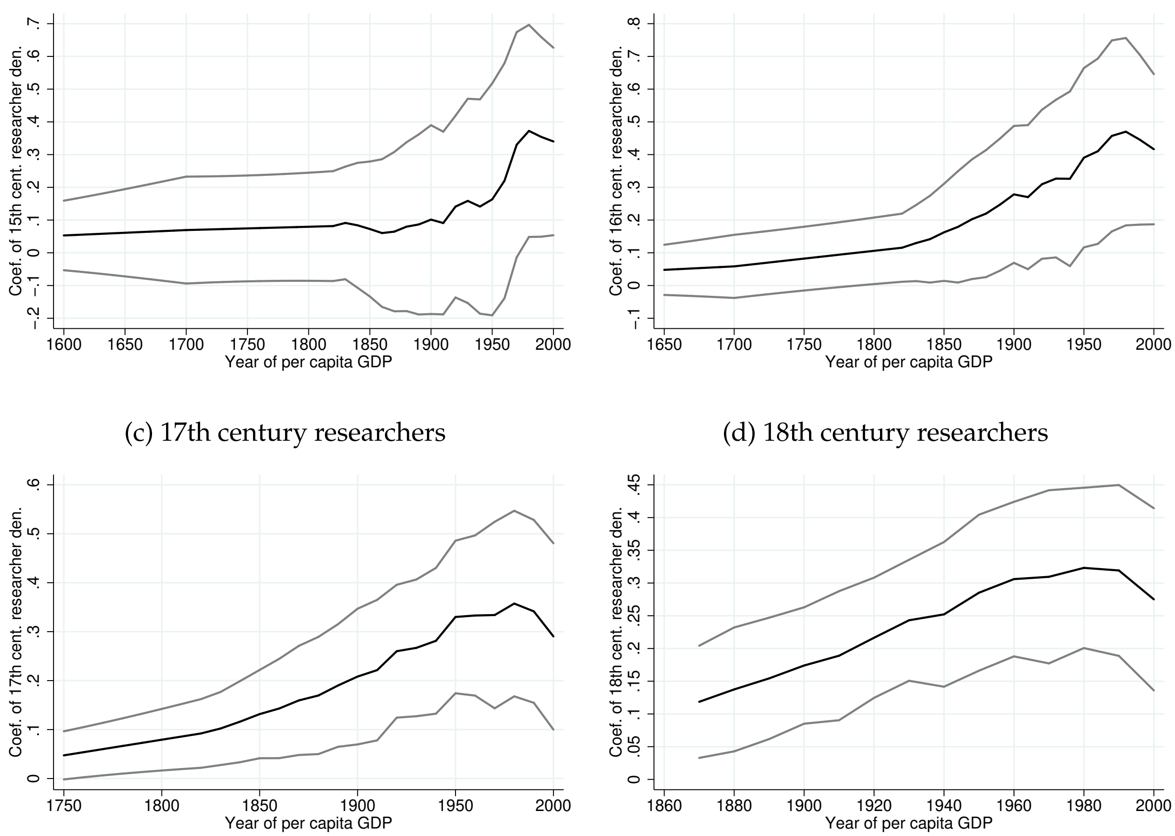
In our baseline cross-sectional estimates we used per capita GDP of 2005 as dependent variable. How do the estimated coefficients change when using other years for the left hand side GDP level? The answer to this question turned out to be a mayor result on its own. First we ran regressions analogous to those in table 4 using distinct dates for the dependent variable:

$$\begin{aligned}\text{Log p.c. GDP}_t \quad = \quad & \text{const.}_t + \lambda_t^i \text{ Researchers p.c. in the i}^{\text{th}} \text{ cent.} \\ & + \gamma_t \text{ Log p.c. GDP}_{i^{\text{th}} \text{ cent. end}} \\ & + \mu_t \text{Maddison Dummy}_{i^{\text{th}} \text{ cent. end}} + \text{residual}_t \qquad (7)\end{aligned}$$

We plotted the estimated coefficient associated with research intensity λ_t^i against the year of the dependent variable in panels (a) to (d) of figure 22. In an attempt to rule out reverse causality, we restricted our analysis to the impact of research intensity on income levels starting 100 years after the end of the century where the researchers were born.

Interestingly, in all graphs we have a positive slope, indicating that the estimated elasticity increases as time passes—e.g. the number of researchers per capita of the 16th century is more important for explaining the GDPs of 2000 than explaining GDPs of, say, 1700. And the same counts for research intensities of other centuries.

Figure 22: Impact of 15th to 18th century per capita researchers on latter income



(a) 15th century researchers

(b) 16th century researchers

(c) 17th century researchers

(d) 18th century researchers

Notes: Estimated coefficient associated with research intensity are obtained from regressions showed in equation 7. All regressions include a constant, dummies for East Asia, Eastern Europe and the Anglosphere, the initial Log GDP p.c. and the Log per capita researchers born in a certain century. *Sources*: See table 4.

Since this is a striking fact, one may ask how robust this finding really is. To the extent that GDP levels show some degree of persistence, one might think that the initial GDP level one is controlling for explains more of the dependent variable of (7) when the time distance between the dependent variable and the control is smaller, while, when this distance is larger, any other regressor which is somehow correlated with the dependent variable will gain importance. But this is not the case Time-slopes do not vary substantially, as presented in table 8.

Fact 19. *The more time passes between income and an earlier historical researcher density, the stronger the cross-country relationship among them.*

What explains fact 19? We interpret it as the conjunction of two previously noted facts: persistent researchers per capita across countries (fact 10) and the rising impact of researchers per capita on economic growth rates (fact 13).

Omitting the components which are homogenous across countries in order to

Table 8: Time varying elasticity of per capita GDP with respect to past per capita researchers

Dependent variable in panels a) and b): λ_t^i of equation (7).

a) Controlling for initial GDP

	15th cent.	16th cent.	17th cent.	18th cent.	19th cent.
Slope	0.000519***	0.00124***	0.00135***	0.00160***	0.00377*
Observations	41	36	26	14	4

b) Not controlling for initial GDP

	15th cent.	16th cent.	17th cent.	18th cent.	19th cent.
Slope	0.000549***	0.000660***	0.00127***	0.00101***	0.00197
Observations	41	36	26	14	4

Notes: See text *Sources*: See table 4.

center our attention on international differences and using the common log approximation of growth rates, we can express a cross country vector of economic growth as

$$y_t - y_{t-1} \approx r_t\beta_t - \gamma_t y_{t-1}, \qquad \gamma_t > 0, \beta_t > 0, \tag{8}$$

where y_t stands for log GDP per capita, r_t is the log researcher density. According to fact 13 we have

$$\frac{\partial \beta_t}{\partial t} > 0$$

We might rewrite (8) as

$$y_t \approx r_t\beta_t + \rho_t y_{t-1}, \qquad \rho_t = 1 - \gamma_t \tag{9}$$

Additionally, we have persistence of researchers per capita (fact 10). Ignoring a trend and any stochastic component—for simplicity and without loss of generality—, we have

$$r_t = \phi_t r_{t-1}, \qquad 0 \leq \phi_t \leq 1 \tag{10}$$

Substituting (10) into (9),

$$y_t = \rho_t y_{t-1} + \phi_t\beta_t r_{t-1}, \tag{11}$$

provides an explanation for fact 12: as r_t shows persistence, one will encounter a significative and positive correlation between current income levels and *past* levels of research density.

Now we can turn as well to an explanation of fact 19. In figure 22 we had shown that the effect that research density of a particular century had on GDP *gained force* as time passed, e.g. researchers per capita of the 16th century are more important explaining GDP levels of 2005 than GDP levels of 1800.

Lagging (11) we obtain

$$y_{t-1} = \rho_{t-1} y_{t-2} + \phi_{t-1} \beta_{t-1} r_{t-2}. \tag{12}$$

Then, substituting (12) for y_{t-1} in 11 yields

$$y_t = \rho_t \rho_{t-1} y_{t-2} + (\rho_t \phi_{t-1} \beta_{t-1} + \phi_t \phi_{t-1} \beta_t) r_{t-2}. \tag{13}$$

For fact 19 to hold we need the coefficient accompanying r_{t-2} in (13) to be larger than the coefficient accompanying r_{t-2} in (12), i.e.:

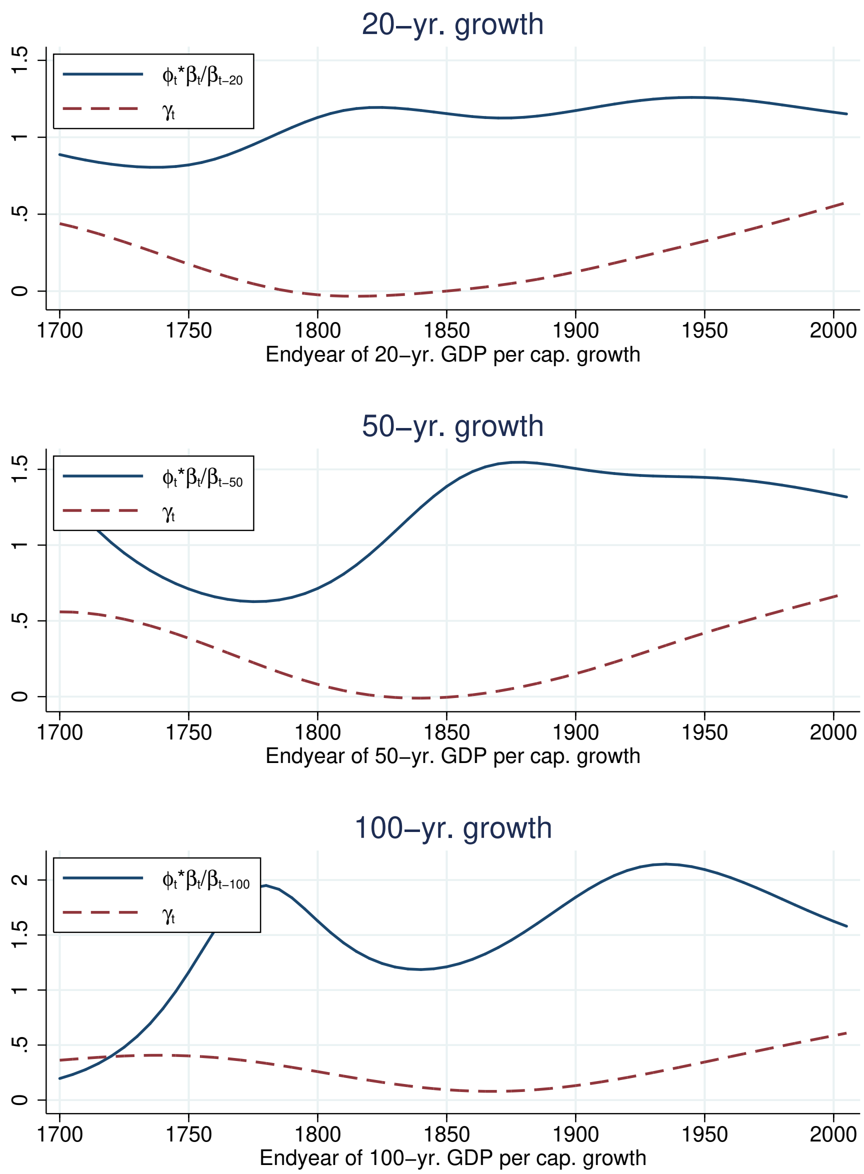
$$\begin{aligned} (\rho_t \phi_{t-1} \beta_{t-1} + \phi_t \phi_{t-1} \beta_t) &> \phi_{t-1} \beta_{t-1} \\ \phi_t \frac{\beta_t}{\beta_{t-1}} &> \gamma_t. \end{aligned} \tag{14}$$

In words, for the research density to gain impact over time, we need the persistence of researchers per capita (ϕ_t) times the growth factor of its impact on GDP (β_t / β_{t-1}) to be greater than the convergence parameter (γ_t).

Is this condition likely to hold? First, if the density of researchers is not persistent at all ($\phi_t = 0$), then (14) cannot hold, i.e. $\phi_t > 0$ is a necessary condition. But we already know that, in practice, ϕ_t is relatively large even in long time intervals, and that β_t has experienced an increasing trend since 1800.

To test if condition (14) holds numerically we plot both $\phi_t \frac{\beta_t}{\beta_{t-1}}$ and γ_t against time in figure 23 for periods of 20, 50 and 100 years of economic growth, using the coefficients obtained for γ_t and β_t from our cross-sectional rolling regressions and ϕ_t obtained from a linear OLS regression of the log number of researchers with respect to its lagged value (and a constant). The estimates depend on the time interval studied, but in all figures we see that in general the condition holds through modern history.

Figure 23: Parameters $\phi_t \frac{\beta_t}{\beta_{t-1}}$ and γ_t of equation (14) for 20-years, 50-years and 100-years economic growth rates



Notes: ϕ_t, γ_t, and β_t correspond to the HP-filtered ($\lambda = 5000$) series of the parameters obtained from the rolling cross-sectional regression.

4 Conclusions

On February 12, 2016, a team of over 1000 physicists published that an observatory in Louisiana had—after measuring distortions of time-space caused by two black holes merging into one—for the first time in history been able to confirm the existence of gravitational waves (Abbott et al., 2016). This happened exactly 100 years after Albert Einstein predicted their existence. So, how much time may it still take for commercial uses of such insights to emerge?

By the same token, when do we expect to see the economic fruits emerge out of the research conducted during Renaissance? Immediately after? 100 years after? 5 centuries after? This is hard to answer. "If inventions were dated according to the first time they occurred to anyone, rather than the first time they were actual constructed, this period [, the Renaissance,] may be regarded just as creative as the Industrial Revolution." (Mokyr, 1990, p. 58).

However, we know that countries where the Renaissance took place are still among those with the highest numbers of researchers per capita. We therefore expect the population of these countries to have been continuously irrigated with new insights about nature, and we expect there to have been people in these societies that have understood these insights and may eventually have used them for their economic benefit. It is not hard to imagine a first self-sustained nuclear reactor appearing in Berlin, Paris, Rome, or Wien instead of Chicago. But does the same count for, say, Caracas? Didn't Venezuelans—which by 1940 had a higher income per capita than Austria, France, and Italy (Maddison, 2010)—lack, above all, the critical amount of engineers and physicists for such achievement? Yes, we believe. And the reasons behind such uneven endowment of high-qualified human capital lie, it seems, in the centurial history of research.

A selected number of economic historians—Joel Mokyr, David Landes, Margaret Jacob, and Jack Goldstone, among them—have put the development of science and research at the center of the history of economic growth. But their "pumpkin books"—as McCloskey (2010) refers to them, because they are big like a "prize-winning pumpkin"—are packed with a detailed historiographical narrative, whose actual economic impact is not easy to compare with the empirical economic literature. Our work instead provides, so to say, a "pumpkin pill"—a concise, assessable, and easy digestible quantification of the "pumpkin narrative". Yet, at the same time, it unveils a number of facts beyond the message of these authors.

How much did we learn from ingesting this pill? First, we can clearly state that research has been mostly exogenous to income. Research grew at modern rates before modern economic growth began—namely during the Renaissance and during the century before 1820. The current number of researchers per capita was determined for most countries before income exploded. And we showed how significant and long-lasting the changes of the research atmosphere in Europe during the Reformation and Counter-Reformation were, a period to which economic historians have perhaps not paid the attention it deserves.

Second, we learned that placing research at the heart of economic development *and* international income inequality, as done by Kuznets, Mokyr and a number of growth theorist, is consistent with the data: research predicts international differences of long-term economic growth, and so more than the main factors proposed by the empirical literature: schooling and executive constraints.

The R^2 of the cross-country regression of 2005 log income with respect to the 19th century researcher density alone is about 0.7. Hence, we may venture to conclude that about 70% of current income inequality is due to the history of research. However, this will be the case only if the explanatory variable is exogenous, highly relevant, and uncorrelated with other relevant factors. As we saw, exogeneity seems to be the case indeed, since historical accidents exogenous to income have been the primary forces determining the geographical distribution of researchers. Nonetheless, model uncertainty in general and, in particular, the omission in our study of some alternative explanations of industrialization, such as the development of trade, the rise of the consumer society, and the role of manufacturing-based industrialization, threatens the validity of such a conclusion, if these factors are shown to be collinear with the researcher densities. Indeed, successful countries did not only invest heavily in research, but also bet on manufactures and global markets. It is questionable, though, whether Germany, for instance, would have become such a large exporter of a complex array of manufactures without its history of research, or whether the East Asian miracle would have been possible without heavy research efforts.

Lastly, we learned that there has been a dynamic impact of research on economic growth and have a quantitative estimate of it. From the times of Francis Bacon—who in the early 17th century was already convinced of the role that research had for economic growth—to our days the impact, though volatile, has been consistently on the rise. Today, for a country which manages to double its number of researchers per capita in a few decades (as was done by Japan and others), per capita GDP may be expected to grow at almost 1% more during the next twenty years.

Beside the main results, this work also hinted to specific conjectures that need more study to be conclusive and found empirical regularities with no obvious explanation. What are researcher densities precisely proxying for? In other words, what is the specific mechanism through which societies with more researchers end up having more economic growth? Why, if income and population growth have been shown to play a minor role, have researchers worldwide grown at a constant rate since the early 18th century? Could it be that research has its own endogenous balanced growth path, e.g. a combination of new research horizons building on previous research but limited by human rational capabilities? What has driven some societies to produce more research than others? And is the relative constancy of economic growth rates in the advanced nations the result of almost constant growth of researchers worldwide? These are just a few of many questions still to be answered.

References

Abbott, B. P. et al. 2016. "Observation of Gravitational Waves from a Binary Black Hole Merger." *Physical Review Letters*, 116(6): 1–16.

Acemoglu, Daron, and Simon Johnson. 2005. "Unbundling Institutions." *Journal of Political Economy*, 113(5): 949–95.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369–1401.

Acemoglu, Daron, and James Robinson. 2013. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. London: Profile Books, 1st edition.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–352.

Allen, Robert C. 2003. "Progress and Poverty in Early Modern Europe." *The Economic History Review*, 56(3): 403–443.

Allen, Robert C. 2009. *The British Industrial Revolution in Global Perspective*. Cambridge, UK: Cambridge University Press, 1st edition.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, 609–626, Princeton: Princeton University Press.

Bénabou, Roland, Davide Ticchi, and Andrea Vindigni. 2015a. "Religion and Innovation." *American Economic Review*, 105(5): 346–51.

Bénabou, Roland, Davide Ticchi, and Andrea Vindigni. 2015b. "Forbidden Fruits: The Political Economy of Science, Religion, and Growth." Working Paper 21105, National Bureau of Economic Research.

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

Biraben, Jean Noël. 1980. "An Essay Concerning Mankind's Demographic Evolution." *Journal of Human Evolution*, 9(8): 655–663.

Bolt, Jutta, and Jan Luiten Zanden. 2014. "The Maddison Project: collaborative research on historical national accounts." *Economic History Review*, 67(3): 627–651.

Broadberry, Stephen. 2013. "Accounting for The Great Divergence." Economic History Working Paper No: 184/2013, London School of Economics and Political Science, Department of Economic History.

Clark, Gregory. 2012. "A Review Essay on The Enlightened Economy: An Economic History of Britain 1700–1850 by Joel Mokyr." *Journal of Economic Literature*, 50(1): 85–95.

Comin, Diego, William Easterly, and Erick Gong. 2010. "Was the Wealth of Nations Determined in 1000 BC?" *American Economic Journal: Macroeconomics*, 2(3): 65–97.

Crystal, David. 2003. *The Cambridge Encyclopedia of the English Language*. Cambridge, UK: Cambridge University Press, 2nd edition.

Gascoigne, Robert M. 1984. *A Historical Catalogue of Scientists and Scientific Books: From The Earliest Times to The Close of The Nineteenth Century*. New York; London: Garland.

Gascoigne, Robert M. 1992. "The Historical Demography of the Scientific Community, 1450-1900." *Social Studies of Science*, 22(3): 545–573.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9(3): 271–303.

Goldstone, Jack A. 2009. *Why Europe? The Rise of the West in World History, 1500-1850*. New York: McGraw-Hill.

Griffith, Rachel, Stephen Redding, and John Van Reenen. 2004. "Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries." *Review of Economics and Statistics*, 86(4): 883–895.

Grossman, Gene M., and Elhanan Helpman. 1993. *Innovation and Growth in the Global Economy*. Cambridge, Massachusetts: The MIT Press, reprint edition edition.

Hall, Bronwyn H., Jacques Mairesse, and Pierre Mohnen. 2010. "Measuring the Returns to R&D." In Nathan Rosenberg, and Bronwyn H. Hall, eds., *Handbook of the Economics of Innovation*, Vol. 2, 1033–1082, Amsterdam: Elsevier.

Jacob, Margaret C. 1997. *Scientific Culture and the Making of the Industrial West*. New York: Oxford University Press.

Jones, Benjamin, E. J. Reedy, and Bruce A. Weinberg. 2014. "Age and Scientific Genius." Working Paper 19866, National Bureau of Economic Research.

Jones, Charles I. 2005. "Growth and ideas." In Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, 1063–1111, Amsterdam: Elsevier.

Jones, Charles I. 2015. "The facts of economic growth."Technical report, National Bureau of Economic Research.

Kamen, Henry. 2014. *The Spanish Inquisition: A Historical Revision*. Yale: Yale University Press.

Knecht, Robert J. 2002. *The French Religious Wars 1562-1598*. Oxford: Osprey Publishing.

Kuznets, Simon. 1966. *Modern Economic Growth*. New Haven: Yale University Press.

Landes, David S. 2003. *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge, UK: Cambridge University Press, 2nd edition.

Maddison, Angus. 1991. *Dynamic Forces in Capitalist Development: A Long-Run Comparative View*. New York: Oxford University Press.

Maddison, Angus. 1997. "Economic Policy and Perormance in Capitalistic Europe." In Louis Emmerij, ed., *Economic and Social Development Into the XXI Century*, Washington DC: Inter-American Development Bank.

Maddison, Angus. 2010. "Statistics on World Population, GDP and per capita GDP, 1-2008 AD." *Historical Statistics*.

Madsen, Jakob B. 2013. "Human Accomplishment and the British Industrial Revolution: The Role of Great Scientists and Education in British Growth since 1270." working paper, mimeo.

Madsen, Jakob B., and Fabrice Murtin. 2015. "The Mechanics of Economic Development in Britain since 1270: The Role of Great Scientists and Education." Macro Workshop, Paris School of Economics, March.

Madsen, Jakob B., and Eric Yan. 2013. "The first Great Divergence and the evolution of cross-country income inequality during the last millennium: the role of institutions and culture." *Applied Economics*, 45(33): 4641–4650.

Marshall, Monty G., Ted R. Gurr, and Keith Jaggers. 2013. "Polity IV project: Political regime characteristics and transitions, 1800-2013."Technical Report Dataset Users' Manual, Center for Systemic Peace.

McCloskey, Deirdre. 2010. "Review of Mokyr's The Enlightened Economy." January.

McEvedy, Colin, Richard Jones, and others. 1978. *Atlas of World Population History*. Harmondsworth, Middlesex, UK: Penguin.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. London: Oxford University Press.

Mokyr, Joel. 2000. "The Industrial Revolution and the Netherlands: Why did it not happen?" *De Economist*, 148(4): 503–520.

Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton: Princeton University Press.

Mokyr, Joel. 2005. "The Intellectual Origins of Modern Economic Growth." *The Journal of Economic History*, 65(02): 285–351.

Mokyr, Joel. 2007. "The Market for Ideas and the Origins of Economic Growth in Eighteenth Century Europe." *Tijdschrift voor Sociale en Economische geschiedenis*, 4(1): , p. 3.

Mokyr, Joel. 2009. *The Enlightened Economy: An Economic History of Britain 1700-1850*. Yale: Yale University Press.

Mokyr, Joel. 2016. *A Culture of Growth: The Origins of the Modern Economy*. Princeton: Princeton University Press.

Murray, Charles. 2003. *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950*. New York: Harper Collins.

Nunn, Nathan. 2014. "Historical Development." In Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 2, 347–402, Amsterdam: Elsevier.

Ogilvie, Sheilagh, and A.W. Carus. 2014. "Institutions and Economic Growth in Historical Perspective." In Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 2, 403–513, Amsterdam: Elsevier.

Rogers, Kara., ed., 2010. *The 100 Most Influential Scientists of All Time*. The Britannica Guide to the World's Most Influential People, New York: Britannica Educational Publishing.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71–S102.

Simmons, John G. 2000. *The scientific 100: A ranking of the most influential scientists, past and present.*: Citadel Press.

Spolaore, Enrico, and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51(2): 325–69.

Spolaore, Enrico, and Romain Wacziarg. 2014. "Long-Term Barriers to Economic Development." In Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 2, 121–176, Amsterdam: Elsevier.

Squicciarini, Mara P., and Nico Voigtländer. 2014. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." Working Paper 20219, National Bureau of Economic Research.

Tarrant, N. 2014. "Censoring Science in Sixteenth-Century Italy: Recent (and Not-So-Recent) Research." *History of Science*, 52(1): 1–27.

Temin, Peter. 2014. "Cambridge History of "Capitalism"." Working Paper 20658, National Bureau of Economic Research.

Tetlock, Philip Eyrikson, and Geoffrey Parker. 2006. *Unmaking the West: "what-if" Scenarios that Rewrite World History*. Ann Arbor, Michigan: University of Michigan Press.

Tiner, John Hudson. 2000. *100 Scientists Who Shaped World History*.: Turtleback Books Distributed by Demco Media.

United-Nations. 1999. "The World at Six Billion."Technical Report ESA/P/WP.154, United Nations' Population Division Department of Economic and Social Affairs.

Vidal-Robert, Jordi. 2013. "War and Inquisition: Repression in Early Modern Spain." CAGE Online Working Paper Series 119, Competitive Advantage in the Global Economy (CAGE).

Vidal-Robert, Jordi. 2014. "Long-run effects of the Spanish Inquisition." CAGE Online Working Paper Series 192, Competitive Advantage in the Global Economy (CAGE).

World Bank, The. 2016. "World Development Indicators." February.

Appendix A Data appendix

A.1 Construction of the uncorrected database

The raw, unadjusted, database was constructed between October 2014 and February 2015 out of queries from DBpedia (dbpedia.org/), a repository containing structured metadata gathered from the English version of Wikipedia. The definition of researcher we employed corresponds to a person engaged in work aimed at increasing the stock of what one might call—following Kuznets (1966)—"useful knowledge", including all branches of theoretical and applied science, mathematics, and philosophy, but excluding arts and religious studies. Trying to match this definition, we first retrieved metadata from profiles of people tagged in DBpedia as "scientist", "philosopher", "inventor", or "intellectual".[30] Some of these persons were rather artists, religious figures, politicians, or national heroes, so that we kept only those with a research-related "shortDescription"—i.e. one containing words such as "researcher", "astronomer", "biologist", and the like.[31] Subsequently we performed further extensive inspection of the data for two reasons. First, since the metadata of Wikipedia is constructed out of an ample communitarian effort, we expected an important part of the data to be erroneous.[32] To our surprise, however, the appearance of vandalized articles was almost nil and the quality of the data proved to be high, since most of our conclusions changed little after months of intensive data cleaning. Second, as our goal has been to perform cross-country analysis, historical places of birth needed to be matched with contemporary geographical limits. Here our criteria was to assign nationalities to people born before modern limits according to their ethnicity or, in cases we find it more appropriate, the university where they taught for most of their life.[33] It is worth noting that migration is often not taken into account in the national categorization of the data: an Hungarian-born young scientist that migrated to the US during World War II, for instance, might be assigned to Hungary despite having been active mostly in the US.

We were particularly meticulous revising the data of people born before 1900, as data going back in time gets scarcer and results based on them more sensitive to sampling errors. Up to those born

[30] We started only with scientists first but noticed that further "types" were needed to include some ancient researchers which were tagged as philosophers, important contributors to engineering which were tagged as inventors, and other scientists labeled as intellectuals but not considered elsewhere. Miguel Servet, for instance, the first European who correctly described the function of pulmonary circulation, was only to be found in the category "intellectual".

[31] Albert Einstein, for instance, is kept because its short description is "Physicist", while Ernesto Guevara—who is strangely also tagged as "Scientist"—is dropped, as its short description is "Argentine-born Marxist, politician, and leader of Cuban and internationalist guerrillas". This filter even dropped Benjamin Franklin, who's short description is "American printer, writer, politician". In order to avoid dropping very important researchers we compared our data with the 100 most influential scientists of all time of Britannica Encyclopedia [Rogers (2010)] and two similar lists [Simmons (2000), Tiner (2000)]. All but 13 out of 240 scientists of these lists were in the filtered database. They were naturally added to the definitive dataset.

[32] Wikipedia provides a review of third-party studies assessing their data reliability in http://en.wikipedia.org/wiki/Reliability_of_Wikipedia

[33] both ancient Greeks scattered around the Mediterranean and Byzantine Greeks were assigned to Greece; Indians born in current day Pakistan or Bangladesh were assigned to India; Transylvanians were assigned to Hungary or Romania according to their ethnicity; Italians living in colonies of the Republic of Venice were assigned to Italy; Turks from around the Ottoman empire were assigned to Turkey while Arabs were assigned to current day nations; ethnical Germans from all over Europe were assigned to Germany; people from Austria-Hungary were assigned to countries where one find their ethnicity today; Bosnians, Croats, Macedonians, Montenegrins, Serbs, and Slovenians were assigned to Yugoslavia; Slovaks and Czechs were assigned to Czechoslovakia; English, French or Dutch colonizers born in Africa, Asia, or the Americas but which returned to Europe were assigned to their European precedence country; some Armenians were included into Turkey despite later genocide, while some other were assigned to Azerbaijan.

in the 18th we were able to inspect manually their birthcountries and the general validity of their metadata, but due to the scale of the dataset, we had to restrict ourselves to selective checking of researchers born in the 19th century (just focusing on countries with small numbers of researchers) and relied exclusively on programming filters for those born in the 20th century.

One difficult task was to determine to which extent religious scholars were to be considered researchers or not. The clergy in Europe and Asia alike was responsible for the safeguarding of countless ancient scripts, including medical studies, philosophical writings and historical resources. Also, during periods such as Arab Al-Andalus or the European Renaissance, people engaged in religious studies were particularly prone to research other areas of scientific interest, so that the presence of influential religious figures may well have been correlated with higher degrees of philosophical and scientific sophistication. Our approach, however, was to keep religious scholars out of the data if they did not arguably made any substantial contribution to areas of knowledge beyond religion itself.

The resulting dataset contains more than 40,000 researchers distributed in time and research categories as presented in table A1.[34] One can note that 28% of researchers born prior to the year 1500 were tagged as philosophers, a number that falls to less than 10% since the 17th century. As noted above, it was important to include this category, since people such as Thales or Pythagoras, for instance, appear as philosophers but not as mathematicians. While it might be that the importance of philosophers in the generation of "useful knowledge" has declined over time as science has gradually drifted apart from philosophy, we decided to keep them uninterruptedly to avoid an artificial brake in the time series and also because it is hard to determine when—if indeed—philosophy ended to be a significant input to technological advances.

Table A1: Researchers by birthdate

	<1500	15th c.	16th c.	17th c.	18th c.	19th c.	20th c.
Number of researchers	334	138	448	702	2944	12279	23803
Scientists[1] and mathematicians	63%	61%	73%	78%	78%	80%	84%
Applied scientists[2] and inventors	11%	12%	19%	16%	20%	21%	16%
Philosophers	28%	12%	10%	9%	5%	3%	4%
Other scholars[3]	13%	30%	16%	11%	8%	6%	8%
Res. with identified birthplace	310	125	430	675	2862	11603	21742
Percentage	0.93	0.91	0.96	0.96	0.97	0.94	0.91

Notes: [1] denotes natural and social scientists. [2] denotes physicians and engineers. [3] denotes researchers with ambiguous descriptions such as "Nobel laureate", "Italian humanist", or "German professor". Because some researchers were active in more than one field, the percentages sum up to more than one. *Source*: Own computations based on DBpedia.

From all researchers, about 35,000—roughly 86% of the whole database—have identified birthplaces and can thus be used in cross sectional analysis. This share increases to almost 100% for researchers born prior to the 19th century as a result of our one by one checking of those researchers.

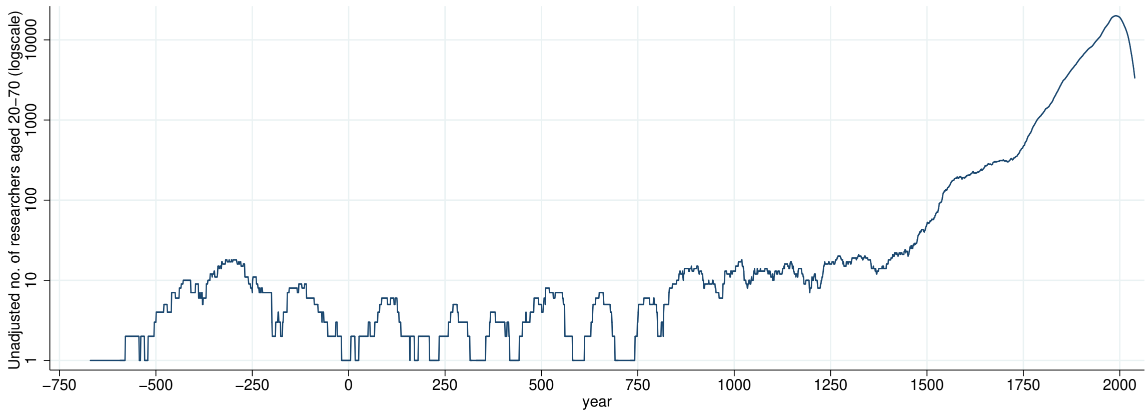
Prior to the 15th century the median number of famous researchers per country is low —even in countries who do have research activity the median number per century is just one at that point—, hence one must abdicate from using cross country indicators that go that far in the past.

The table also shows how the number of researchers increases exponentially over time—and so does the reliability of the cross country indicators as well.

Figure 24 shows the number of researchers aged 20 to 70 (including deceased ones) through time

[34]As some researchers might appear in more than one category, percentages added vertically in the table do add up to more than 100%.

Figure 24: Researchers aged 20-70 worldwide (not adjusted for geographical biases)



Source: Own computation based on DBpedia.

A.2 Comparison to alternative sources

The world aggregate and the takeoff of research. Figure 25 presents a comparison between the aggregate series based on DBpedia and on Gascoigne (1984)—to the best of our knowledge, the largest catalogue of scientist available for early modern times. It shows that the main messages of DBpedia, namely that worldwide researchers exploded between the 15th and 16th centuries, grew at a lower rate during the 17th century, and exhibited steaty-state growth afterwards, is mirrored by Gascoigne's data. Gascoigne (1992) shows how his series presented in figure 25, which is based on a list of 12,338 scientists born between 1410 and 1860, also closely mimics the evolution of other sources, such as that Dictionary of Scientific Biographies, which includes 3062 scientists for that period, and his Chronology of the History of Science, which contains just 1000 scientists. His analysis has important implications. It tells us that measures of research output are highly correlated. Therefore, conclusions drawn from DBpedia data, which is subject to an arbitrary "entry barrier", are likely to extend to databases with more researchers as well. Smaller datasets will, of course, present higher variance and tend to be less reliable for econometric estimates.

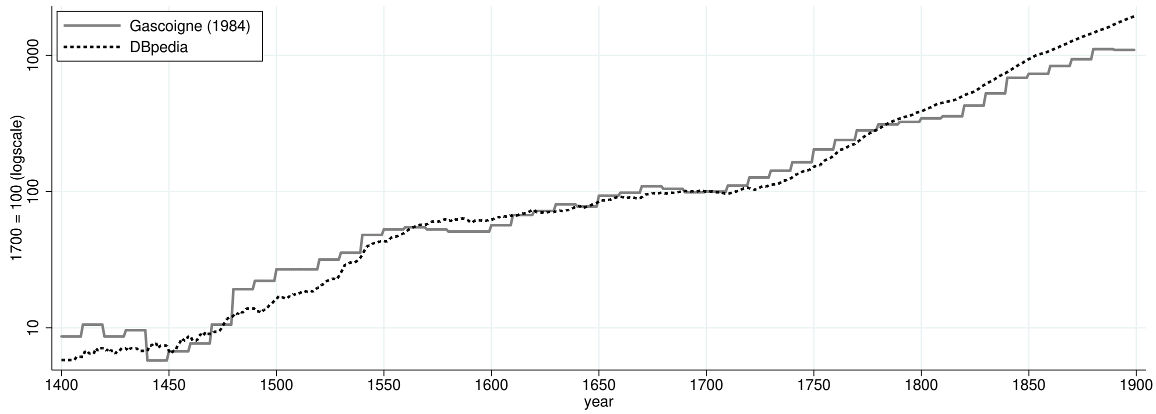
Research within countries. Figure 26 shows the evolution in time of the number of researchers across some Western countries. It presents, again, data of Gascoigne (1984) but also the much more restrictive data of Murray (2003), which is based on about 1000 "great minds" of science previous to 1900, hence representing the most significant figures contained in Gascoigne and DBpedia. When differences emerge between Gascoigne's data and DBpedia, we can see that the reason is that DBpedia tracks more closely the evolution captured by Murray in some cases. In other words, DBpedia data lies somewhere between Gascoigne (1984) and Murray (2003).[35]

Figure 27 compares the number of researchers in a selected group of Western countries with data of Murray (2003). As in figure 5 , Britain dominates the research landscape between 1650 and about 1850, closely followed by France and Germany, while Italy falls behind. Similar developments can be found in figure 28, albeit with an overrepresentation of German scientists, which is due to the "strength of the German tradition of comprehensive biography", as Gascoigne (1992) acknowledges.

Research across countries. Figure 29 shows the cross sectional relationship between DBpedia and Gascoigne (1984). Further graphical comparison of DBpedia with alternative metrics of research can be found in figures 30a, 30b and 30c. Figure 30a plots the number of active researchers per

[35] Gascoigne, 1992 acknowledges some biases in its sampling of researchers. One is the overrepresentation of German scientists due to the "strength of the German tradition of comprehensive biography".

Figure 25: The world aggregate



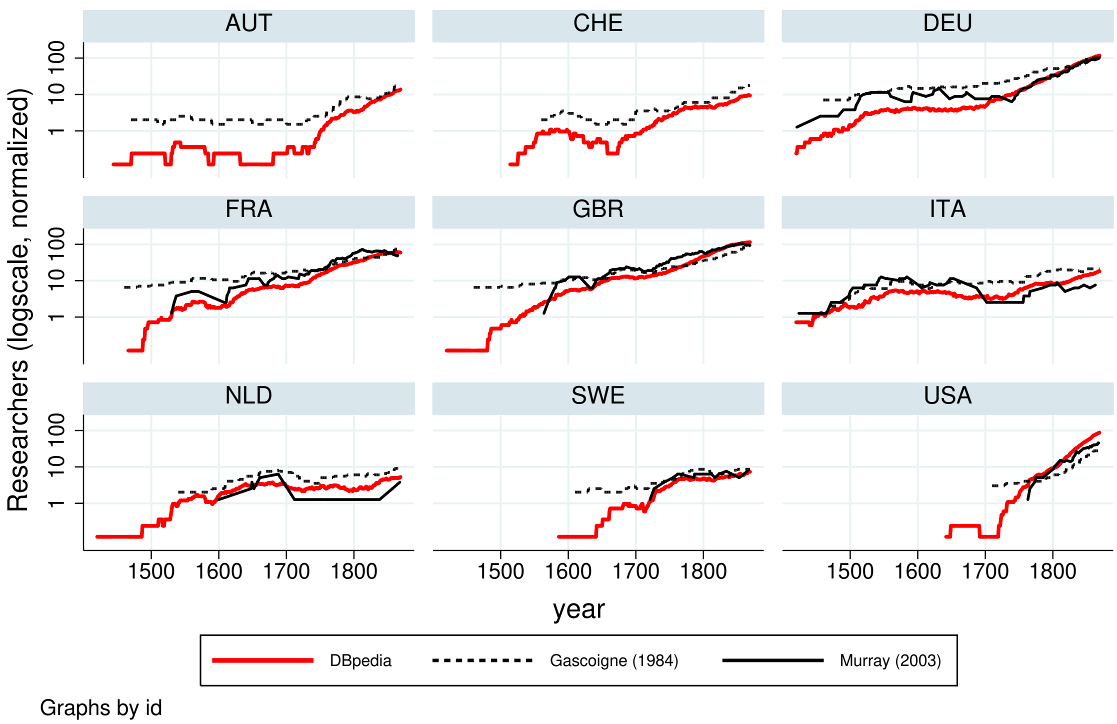
Notes: Gascoigne's data is presented as a three-decades moving average of people born in a decade, as presented in Gascoigne (1992), but lagged 40 years in order to present people with 40 years instead of newly born to-be researchers. DBpedia data corresponds to researchers aged 20 to 70. Both series are normalized to 100 in 1700. *Sources*: Gascoigne (1984) (as in Gascoigne, 1992) and own computation based on DBpedia.

capita in 2005 from DBpedia against the one provided by the UNESCO for the World Bank's World Development Indicators (WDI). We do not expect a perfect fit among both measures. Researchers appearing in Wikipedia must be famous, so that the DBpedia-based metric is an output indicator. On the contrary, researchers appearing in official statistics need not to be associated to any known significant production. Official numbers of researchers are rather an input indicator. Yet, despite differences such as a clear underrepresentation of East Asia in DBpedia, we do observe a relationship with almost unitary elasticity, as would be expected from an unbiased indicator.

A similar comparison is made in figure 30b. It plots for year 1981 the number of researchers per capita from DBpedia against the number of scientific articles per capita provided by the National Science foundation (NSF). Now being both output indicators, the fit is considerably better.

And a comparison with earlier historical densities of researchers is provided in figure 30c. It presents the per capita number of researchers born in the 19th century according to DBpedia against the same number according to Murray (2003), which condenses 34 sources of the history of science. Albeit some differences (Murray's data is based on 10 times less researchers), a marked positive relationship between the two sources can be observed.

Figure 26: Gascoigne (1984), Murray (2003), and DBpedia



Notes: The series are normalized to 100 for the maximum in 1850 (not per country), being therefore comparable across countries. Gascoigne's data is presented as a three-decades moving average of people born in a decade, as presented in Gascoigne (1992), but lagged 40 years in order to present people with 40 years instead of newly born to-be researchers. DBpedia and Murray's data correspond to researchers aged 20 to 70. *Sources*: Gascoigne (1984) (as in Gascoigne, 1992), Murray (2003), and own computation based on DBpedia.

Figure 27: Researchers aged 20-70 in selected countries. Murray's (2013) data.

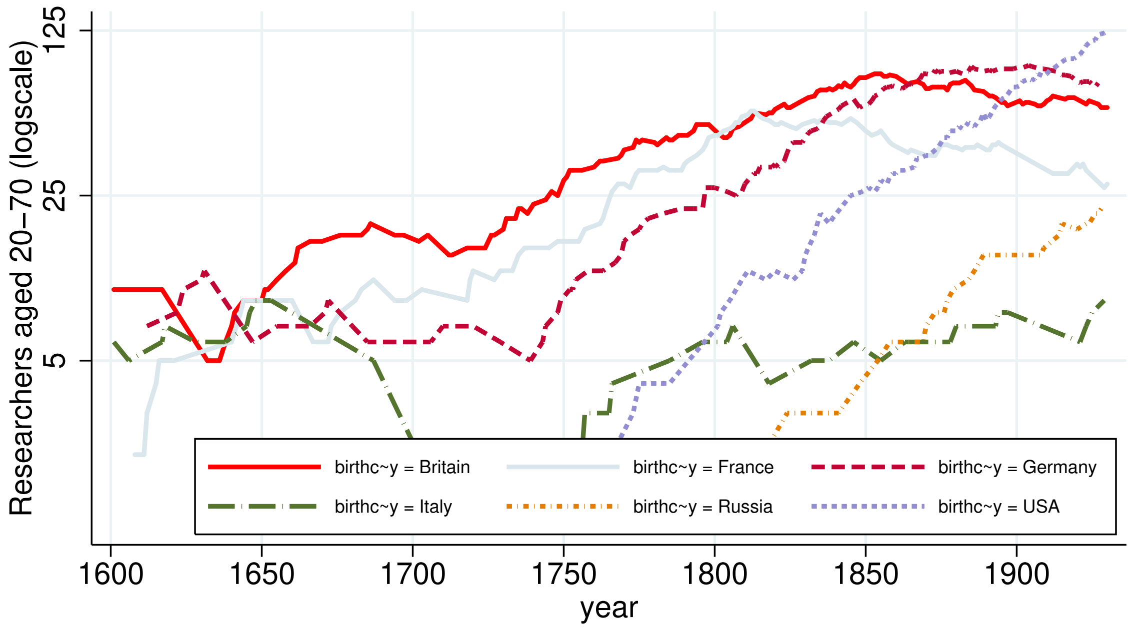
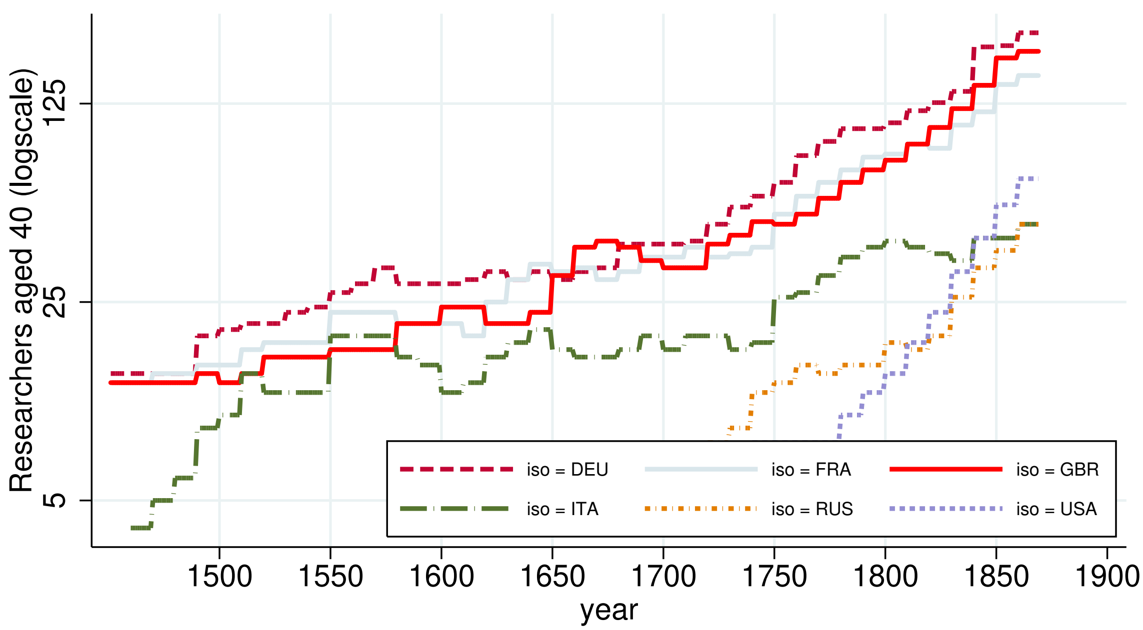


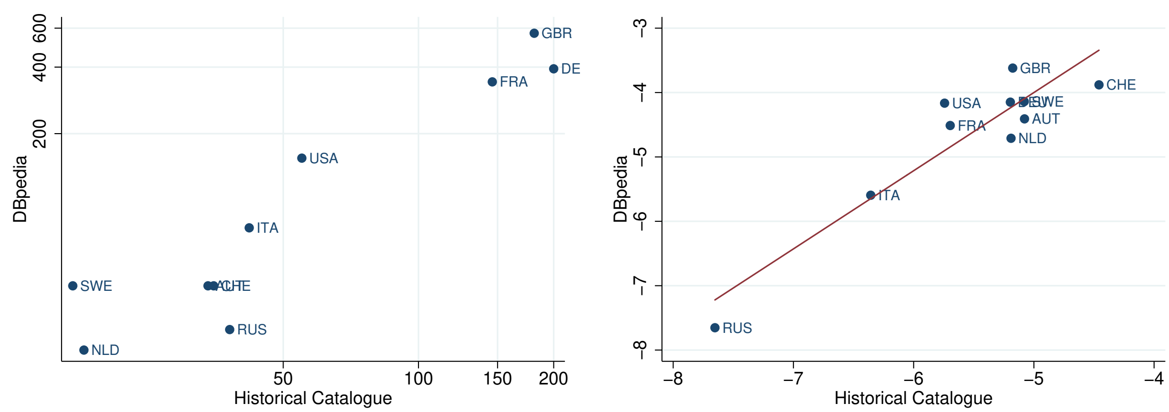
Figure 28: Researchers aged 40 in selected countries. Gascoigne HCS



Notes: See text.

Figure 29: Researchers per capita in 1820. Gascoigne vs. DBpedia

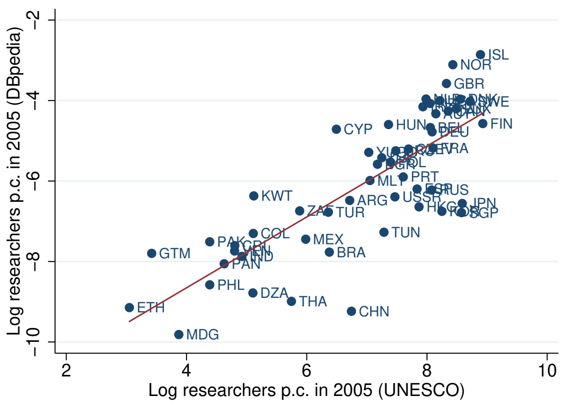
(a) Researchers in 1820 (logscale) (b) Researchers per capita in 1820 (in logs)



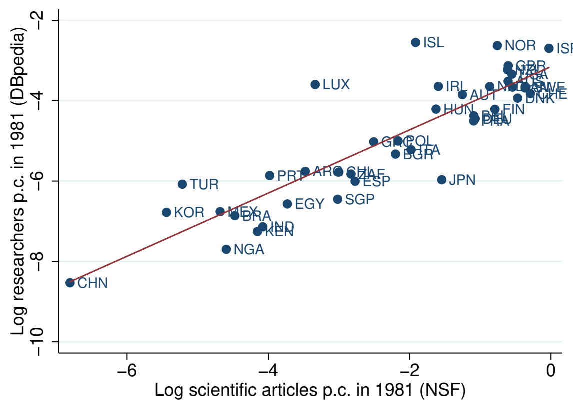
Notes: See text.

Figure 30: Unadjusted DBpedia log researchers per capita vs. . . .

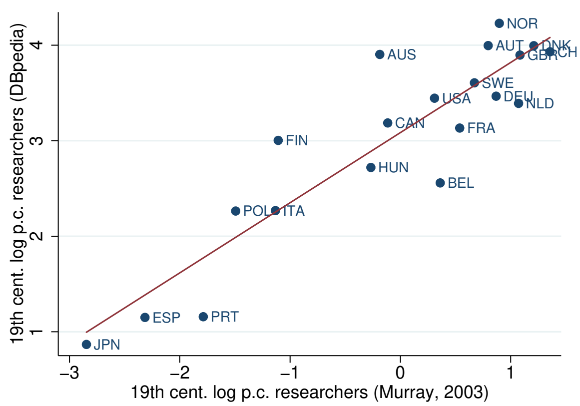
(a) . . . log UNESCO researchers p.c., 2005



(b) . . . log scientific articles, 1981



(c) . . . log scientists p.c., 19th cent.



Notes: Researchers from DBpedia correspond to those aged 20 to 70 in a given year. *Sources*: UNESCO researchers and NSF articles are from World Bank (2016), scientists are from Murray (2003), population is from Maddison (2010), and researchers correspond to own computations out of DBpedia.

A.3 Correction of the database

However, even a perfect fit in figure 30c would *not* mean that the data is unbiased. The work of Murray has indeed been accused of "exaggerated Eurocentrism" (Tetlock and Parker, 2006, p. 5). And, of course, Eurocenstrism—or, more generally, a bias towards high income countries—is a serious threat for studying the relationship between researchers and subsequent economic development when relying on cross-sectional regressions. In such case a positive relationship between income and research could arise solely out of reverse causality, as shown schematically in equation (15):

$$\text{Income today} \uparrow \Rightarrow \text{DBpedia use} \uparrow \Rightarrow \text{researchers in all years} \uparrow \tag{15}$$

Is Wikipedia biased towards higher income countries? It sounds plausible that an average editor of Wikipedia is more likely to write about researchers from his country than about researchers from other parts of the world. Therefore, as our data is based on entries made by editors of the *English* Wikipedia and because internet penetration is stronger in higher income countries, a sampling bias could potentially arise in the form of an overrepresentation of researchers from countries with high income and high shares of English-speaking population.

To inspect the degree of such potential biases we may look at how close DBpedia researchers are tracking other similar measures which are definitively not affected by these potential biases. One such measure should be the official number of per capita researchers which we introduced in figure 30a. Also the number of articles per capita of the NSF presented in figure 30b might be a valid proxy of an unbiased point of comparison.

A statistical comparison between these alternative proxies of research and the DBpedia index is presented in panels A, B and C of table A2. In columns (1) to (4) of panel A the log difference in 4 different years between the researcher density based on DBpedia and the number of articles per capita is regressed with respect to log per capita GDP of 2010, a dummy for East Asia, and a constant. The same is done in columns (5) to (7), with the dependent variable being now the log difference between the researcher density based on DBpedia and the one based on official statistics for three different years. All these regressions are aimed at showing whether there is an overrepresentation of high income countries in DBpedia. But the coefficient is negative or not significantly different from zero. Hence, high income countries are not overrepresented in DBpedia.

A similar comparison is made in panel B, now adding as regressors the intensity of internet users and a set of regional and linguistic dummies. Surprisingly, neither are the Eastern European block and French speaking countries underrepresented, nor is the Anglosphere (USA, UK, Canada, Ireland, Australia, and New Zealand) overrepresented.

It would be desirable to control for the number of editors of Wikipedia in each country or, at least, for the number of English speakers. But, unfortunately, there are no reliable statistic on either of both measures. To the best of our knowledge, the only comparable metric of English speakers is the one provided by the Eurobarometer of the European Commission, covering just some European countries. Panel C presents how this measure relates to the same dependent variable as in the previous panels. Except for column (1), the more English speakers a country has, the more represented are their researchers in Wikipedia.[36] Therefore, table A2 suggest no systematic overrepresentation of high income countries, but a significant underrepresentation of East Asian countries and countries with few English speakers. This means we cannot compare the UK with, say, China or France, unless we perform a correction of these biases.

When performing comparisons among European countries, we will employ correction (1) of the density of researchers, repeated here for convenience, at any point in time:

$$\begin{aligned}\text{Adjusted log researchers p. c.}_t &= \text{Unadjusted log researchers p. c.}_t \\ &\quad + (100 - \%\ \text{English speakers}) \times 0.0125 \end{aligned} \tag{1}$$

where the number 0.0125 corresponds to the average of columns (2) to (7) in panel C of table A2. We will refer to this adjustment as the English-correction.

[36] Using less reliable sources with intercontinental coverage, such as Crystal (2003) or the English Profilency Index, produce almost the same coefficient. Results can be provided upon request.

When making comparison among countries of the entire sample we will employ the alternative, suboptimal (less precise), correction instead. It is based on the residuals of regression (2):

$$\begin{aligned}\text{DBpedia log researchers p. c.}_t &= \text{const.} + \text{NSF log articles p. c.}_t \\ &+ \text{residual}_t, \qquad t = 1986, \ldots, 2000 \end{aligned} \tag{2}$$

These capture the average degree of over or underrepresentation in DBpedia with respect to the density of articles according to the NSF.[37] The adjusted metric, to which we will refer to as NSF-correction, is thus

$$\begin{aligned}\text{Adjusted log researchers p. c.}_t &= \text{Unadjusted log researchers p. c.}_t \\ &- \text{average residual of (2)}\,. \end{aligned} \tag{3}$$

[37]The choice of a metric based on NSF articles instead of a metric based on UNESCO researchers is because the latter offers to few observations.

Table A2: Are there any biases?

A) Is there an income bias?

	ln(DBpedia res. p.c. / NSF art. p.c.)				ln(DBpedia / UNESCO res. p.c.)		
	(1) 1981	(2) 1990	(3) 2000	(4) 2005	(5) 1996	(6) 2000	(7) 2005
ln(GDP/pop) 2010	0.0661	-0.445****	-0.566****	-0.545****	0.139	0.101	0.127
	(0.156)	(0.0894)	(0.0899)	(0.0807)	(0.177)	(0.136)	(0.107)
East Asia dummy	-0.0867	-0.397	-1.070***	-1.339****	-1.591****	-2.013****	-2.472****
	(0.453)	(0.359)	(0.387)	(0.347)	(0.365)	(0.333)	(0.331)
Observations	40	95	96	96	32	41	49
R^2	0.006	0.227	0.347	0.406	0.403	0.493	0.549

B) Are there other biases?

	ln(DBpedia res. p.c. / NSF art. p.c.)				ln(DBpedia / UNESCO res. p.c.)		
	(1) 1981	(2) 1990	(3) 2000	(4) 2005	(5) 1996	(6) 2000	(7) 2005
ln(GDP/pop) 2010	0.225	-0.161	-0.160	-0.335	-0.475	0.0919	0.189
	(0.317)	(0.227)	(0.229)	(0.207)	(0.464)	(0.367)	(0.292)
Internet users in 2013(%)	-0.00969	-0.0154	-0.0219**	-0.0127	0.0195	-0.000621	-0.00304
	(0.0140)	(0.00984)	(0.0101)	(0.00916)	(0.0165)	(0.0153)	(0.0118)
Anglosphere dummy	0.0240	0.190	0.352	0.420	0.656	0.399	0.421
	(0.432)	(0.481)	(0.496)	(0.450)	(0.453)	(0.447)	(0.390)
Francophone dummy	0.338	0.617	0.693	0.618	-0.191	-0.218	-0.334
	(0.576)	(0.562)	(0.670)	(0.607)	(0.622)	(0.541)	(0.484)
Eastern Europe dummy	-0.0364	1.149*	0.734	0.770	-0.327	-0.486	-0.388
	(0.660)	(0.631)	(0.567)	(0.514)	(0.503)	(0.508)	(0.464)
East Asia dummy	-0.0656	-0.334	-1.023***	-1.268****	-1.528****	-2.015****	-2.483****
	(0.488)	(0.361)	(0.387)	(0.351)	(0.384)	(0.358)	(0.347)
Observations	40	95	96	96	32	41	49
R^2	0.030	0.277	0.391	0.435	0.506	0.527	0.578

C) English speakers in Europe

	ln(DBpedia res. p.c. / NSF art. p.c.)				ln(DBpedia / UNESCO res. p.c.)		
	(1) 1981	(2) 1990	(3) 2000	(4) 2005	(5) 1996	(6) 2000	(7) 2005
English speakers (%)	-0.00415	0.00681	0.0102	0.0115*	0.0170***	0.0171**	0.0125**
	(0.00863)	(0.00888)	(0.00745)	(0.00653)	(0.00499)	(0.00706)	(0.00526)
Observations	19	21	21	21	14	16	21
R^2	0.013	0.030	0.091	0.140	0.492	0.294	0.230

Notes: OLS estimates with robust standard errors in parentheses. Significance at the 10%, 5%, 1%, and 0.1% levels are indicated by *, **, ***, and **** respectively. Constants included but omitted in the table. In terms of ISO-3166 countrycodes, the dummies are conformed by the following. East Asia: HKG, MAC, JPN, KOR, PRK, VNM, SGP, THA, IDN, MYS, KHM, and TWN; Anglosphere: USA, GBR, CAN, AUS, IRL, and NZL; Eastern Europe: CZEV, CZE, BGR, POL, EST, LTU, LVA, RUS, UKR, ROM, HUN, BLG, ALB, SVK, SVN, HRV, YUG, BIH, YUGO, USSR, EEC, SRB, ROU; Francophone: FRA LUX MCO BEL MUS AND; *Sources*: Per capita income levels are from the Penn World Tables (PWT) 8.0. Population levels are from Maddison (2010). Internet users per capita are from the WDI. Researchers are from DBpedia and from the WDI.