The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000-1800)*

David de la Croix[†] Frédéric Docquier[‡] Alice Fabre[§] and Robert Stelter[¶]

April 13, 2021

Abstract

We argue that market forces shaped the geographic distribution of upper-tail human capital across Europe during the Middle Ages, and contributed to bolstering universities at the dawn of the Humanistic and Scientific Revolutions. We build a unique database of thousands of scholars from university sources covering all of Europe, construct an index of their ability, and map the academic market in the medieval and early modern periods. We show that scholars tended to concentrate in the best universities (agglomeration), that better scholars were more sensitive to the quality of the university (positive sorting) and migrated over greater distances (positive selection). Agglomeration, selection and sorting patterns testify to a functioning academic market, made possible by the use of a common language (Latin).
Keywords: Upper-Tail Human Capital, Universities, Discrete choice model, Scholars, Publications, Agglomeration.
JEL codes: N33, O15, I25.

---

*We thank the research assistants who have been contributing to this adventure: Basile Burnet, Guillaume Catoire, Valentine Debois, Julie Duchêne, Clémence and Timothée de la Croix, Thomas Einsfeld, Maximilian Ganterer, Hugo Jay, Soraya Karioun, Victoire Michel, Annika Onemichl, Gaia Spolverini, Baptiste Standaert, Mara Vitale, Stefanija Veljanoska, and Pascaline Zwarts. Technical help was obtained from Yves Croissant (the author of the package `mlogit`) and Alain Guillet (who reprogrammed the simulation command for a nested-logit model). The project benefited from detailed discussions with Ran Abramitsky, Thomas Baudin, Simone Bertoli, Matteo Cervellati, Gregory Clark, Zvi Eckstein, Sven Feldman, James Fenske, Martin Fernandez-Sanchez, Mattia Fochesato, Oded Galor, Paola Giuliano, Paula Gobbi, Simone Meraglia, Omer Moav, Simone Moriconi, Luca Pensieroso, Giovanni Peri, Itay Saporta, Yannay Spitzer, Nico Voigtlaender, Romain Wacziarg, and Marlous van Waijenburg, and from comments at the following conferences and seminars: WEHC Boston, ULB Brussels, TSE Toulouse, Uni. Luxembourg, LISER (5th workshop on Economics of Migration), PSE Paris (Conference on Culture, Institutions and Prosperity), NYU Abu Dhabi, Deep Rooted Factors of Growth (Brown), CEPR Economic History Workshop (Tarragona), CEMIR Workshop on Migration Research (Munich), 12th Migration and Development Conference (Madrid), U Bologna (1st workshop on Economic History), IDC (Herzliya), Tel Aviv, Hebrew U (Jerusalem), Lille, UCLA, UC Davis, and Standford U. The first two authors acknowledge financial support from the EOS program of the Flemish (FWO) and French-speaking (FRS-FNRS) communities of Belgium (convention 30784531 on "Winners and Losers from Globalization and Market Integration: Insights from Micro-Data").

[†]IRES/LIDAM, UCLouvain, Belgium; CEPR, London

[‡]LISER, Esch-sur-Alzette, Luxembourg

[§]Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE, France

[¶]University of Basel, Switzerland; Max Planck Institute for Demographic Research, Rostock, Germany

1 Introduction

Both scholars and universities are thought to have played a role in the Rise of the West (Mokyr 2016, Cantoni and Yuchtman 2014). We argue and establish empirically that a functioning academic market in the pre-industrial period was a powerful institution allowing them to operate together, helping universities to harness the potential of upper tail human capital at the dawn of the Humanistic and Scientific Revolutions as well as, to a lesser extent, during the subsequent European primacy. Our results shed light on the importance of medieval roots in fostering scientific output, embedding in individual data the qualitative studies on the subject.

Universities are one of the most original creations of the Western Latin civilization during the Middle Ages, from the 11th century onwards.[1] They came into existence when society recognized that masters and students as a collective (*universitas* means community) had legal rights. Universities are voluntary, interest-based, and self-governed permanent associations (Greif 2006). As highlighted in Rashdall (1895), "such Guilds sprang into existence, like other Guilds, without any express authorisation of King, Pope, Prince, or Prelate. They were spontaneous products of the instinct of association that swept over the towns of Europe in the course of the eleventh and twelfth centuries."

A university was thus originally a guild of either students or masters. Near the end of the 12th century, foreign law students at Bologna formed a union for the purpose of protection from discrimination by the town against foreign residents. At about the same time, teachers in Paris formed a corporation. Universities then began to spread across Europe, either through secession from existing ones (Cambridge from Oxford, Padua from Bologna, Orléans from Paris, etc.), or through creation *ex nihilo*. Some universities were founded from scratch by a higher authority (the University of Naples was arguably the first of this kind), but all followed the guild-like organizational principles of Bologna and Paris. Even at the Imperial Moscow University (established in 1755, charter of 1804), the rector was elected by his peers, not nominated by the emperor.

The European academic world in the medieval and early modern era provides a rich background for identifying location patterns within the upper tail of the skill distribution. The use of Latin helped mobility and, despite the political fragmentation of Europe, medieval universities were recognized for their independence and intellectual unity. The academic market was even formalized via the *licentia ubique docendi* (licence to teach everywhere), granted by the Church to the universities at the end of the thirteenth century, and conferring the right to teach at every university in Europe once a doctoral degree had been awarded (Hermans and Nelissen 2005). Understanding the mobility of academic scholars in that period matters because it potentially influenced the creation of knowledge in the pre-industrial period, as well as technological and

[1] A few notable exceptions outside Europe: the Buddhist university of Nalanda in India, where both students and masters are known to come from distant places (Monroe 2000), and the University of Baghdad, which was destroyed by the Mongol invasion in 1258 CE.

institutional progress. Focusing on a period from 1000 CE to 1800 CE,[2] our paper investigates whether location decisions were associated with distance and with measures of individual and institutional quality. We distinguish three notions of quality. The *human capital of an academic scholar* is built from her/his achievements as seen today in the catalog of world libraries (Worldcat). The *notability of a university* in a given period is built from the human capital of its five best scholars. The *simulated output of a university* is the aggregation of the human capital of all scholars who were predicted to work there in a given period.

Although the economic literature has looked at the characteristics of migrant workers at different periods in history, little is known about the mobility of upper-tail human capital in general, and about the internationalization of medieval and early modern European universities in particular. To tackle this question, we develop a unique database that provides geolocalized information on the origin of thousands of academic scholars, on the location of universities, and on measures of individual human capital and institutional notability. We use it to estimate the effects of distance, the human capital of scholars, the notability of universities, and the attractiveness of European cities on location decisions. More specifically, we test (i) whether academic scholars tended to concentrate in the best universities in medieval Europe (*agglomeration*), (ii) whether those with more human capital were more likely to settle in more prestigious universities and/or in more attractive cities (*positive sorting*),[3] and (iii) whether they were more mobile than others (*positive selection*).[4] We finally use our estimated location choice model to compute the potential gains in the output of universities resulting from the agglomeration, positive sorting, and positive selection of academic scholars.

Our database builds on secondary sources (i.e. books and catalogues recovering information from institutional archives) and biographical dictionaries. It documents the mobility and the human capital of 23,624 academic scholars over the whole period 1000-1800. Their location choice set varies across sub-periods, as new universities were created or disappeared over time. On average, each scholar selected their optimal place of work out of 100 possible locations. Our database includes about two and a half million possible dyads (i.e. scholar–university pairs). By studying the mobility patterns of academic scholars in the medieval and early modern periods, we capture a substantial part of upper-tail human capital. The two other – less numerous – groups were the members of scientific academies that developed in Europe in the 17th century (preceded by the Renaissance academies in Italy in the 16th century), and the scholars making a living at the courts of princes, kings, or bishops.

We estimate the mobility patterns using a multinomial logit model, and several variants ac-

[2] Although the official creation date of the first university (Bologna) is 1088 CE, many universities were active before they were formally recognized (see Section 3).

[3] In its common meaning, sorting is any process of arranging items systematically, and has two common, yet distinct meanings: ordering (arranging items in a sequence ordered according to a criterion) and categorizing (grouping items with similar properties). In the migration literature, it means that individuals with better attributes tend to concentrate in regions where returns are higher.

[4] In biology, positive natural selection is the force that drives the increase in the prevalence of advantageous traits. In our location choice model, we test whether better scholars are less sensitive to the distance from their birthplace.

counting for sample biases, heterogeneous effects, and endogeneity issues. We show that agglomeration forces are at work: the destination choice of academic scholars depended on distance, on the notability of the university, on the size of the city (used as a proxy for its economic development), and on the communal freedom enjoyed by the city (used as a proxy for local democracy). We also find robust evidence that better scholars were less sensitive to distance (positive selection) and more sensitive to the attractiveness of the university (positive sorting). Agglomeration and sorting patterns testify to the existence of a functioning academic market in Europe. Such market forces governed the concentration of upper-tail human capital across Europe and the total production of knowledge. They played an important role when there were few universities. By contrast, selection patterns tended to scatter talent across universities, and hardly influenced the aggregate production of knowledge. Agglomeration and sorting substantially helped universities to create knowledge at the dawn of the Scientific Revolution and during the subsequent European primacy. These effects became negligible later when the number of universities increased.

Our paper speaks to three strands of literature. Firstly, we contribute to the literature on stagnation to growth and on the role of upper-tail human capital. Many authors have searched for the profound causes of the "Rise of the West" (e.g. Landes 1998; Maddison 2007; Galor and Moav 2002;[5] Galor 2011; Mokyr 2010; Mokyr 2016). For most of them, the self-reinforcing dynamics of technological and institutional progress played a key role. In particular, De la Croix, Doepke, and Mokyr (2018) argue that superior institutions for the creation and dissemination of productive knowledge help explain the European advantage in the medieval and early modern periods. The outstanding debate concerns the key forces that made these virtuous circles possible. There are currently no global quantitative analyses of the historical effect of upper-tail human capital on the dynamics leading to the Industrial Revolution. Recent country-level studies include Dowey (2017) for England, Squicciarini and Voigtländer (2015) for France, and Dittmar and Meisenzahl (2019) and Cinnirella and Streb (2017) for Germany. Squicciarini and Voigtländer (2015) show that the number of people who subscribed to Diderot's and d'Alembert's *Grande Encyclopédie* in 18th-century France predicts economic development later on, both at the city and county levels. Dittmar and Meisenzahl (2019) show that German cities that adopted better institutions following the Reformation grew faster and had more people recorded as famous in the German biography database.

There is a debate about whether or not universities facilitated the Scientific Revolution. It is true that the new science developed in the 16th century came into conflict with the traditional Aristotelian approach taught at universities. Still, following Applebaum (2003), 87 percent of the scientists listed in the *Dictionary of Scientific Biography* born between 1450 and 1650 were university educated, and 45 percent of them were employed by universities. Beyond science,

[5] Galor and Moav (2002) explicitly refer to the universities: "Further, unlike the existing literature, investment in human capital increased gradually in the Pre-Industrial Revolution era due to a gradual increase in the representation of individuals who have higher valuation for offspring's quality. (...) In particular, in the Pre-Industrial Revolution era, the increase in the number and size of universities in Europe since the establishment of the first university in Bologna in the eleventh century had significantly outpaced the growth rate of population."

medieval universities may have contributed to the rise of the West through (i) the revival of Roman law, which was better suited to regulating complex economic transactions than the prevailing customary law,[6] (ii) the translation of philosophical and scientific works from Classical Arabic and Greek, (iii) the diffusion of scientific thinking in Europe (e.g. Ockham's parsimony principle, Duns Scotus's logic, or Roger Bacon's empiricism), (iv) the promotion by theologians of cultural norms such as the nuclear family, strict monogamy (De la Croix and Mariani 2015), and the education of children (Thomas Aquinas), and (v) the interest in the natural sciences, reflected in the establishment of botanical gardens next to medical faculties. A recent work by Dittmar (2019) lends credence to the idea of higher productivity of university scholars during the Renaissance. Dittmar computes the real wage of 3,000 Italian professors during the Renaissance from archived payrolls, and shows that the premium of those involved in the new sciences increased after the adoption of the movable-type printing press. To our knowledge, this is the only paper other than ours focusing on university professors and using individual-level data.

Beyond the existence of universities (Cantoni and Yuchtman 2014) and the role of elites (Dittmar and Meisenzahl 2019), we stress what makes them operate better together: the academic market. Higher education institutions and elites are present as soon as a civilization reaches a certain level of sophistication, but European universities were unique as they were bottom-up institutions operating in a continental market without many barriers (common language, political fragmentation of Europe). This allowed scholars to sort and concentrate, increasing thereby the output of the whole academic system.

Secondly, our paper relates to the migration literature in general, and to historical migration in particular. Migration is a selective process, with some individuals choosing to leave their region of birth and others choosing to stay. Who moves and who stays depends on the costs and benefits of migration, which can vary across individuals for both systematic and idiosyncratic reasons. Two salient features of contemporary labor mobility are that well-educated people exhibit a much greater propensity to emigrate than the less educated, and they tend to agglomerate in countries/regions with high rewards to skill (Grogger and Hanson 2011; Beine, Docquier, and Ozden 2011; Kerr et al. 2017; Kerr et al. 2016). The geographic concentration of talent is stronger within the upper tail of the skill distribution and does not necessarily lead to a decline in returns to skills due to agglomeration spillovers. Skill-intensive clusters allow better technology exchanges, deeper labor market specialization, or strong complementarities (Stephan and Levin 2001; Kerr, Kerr, and Lincoln 2015a; Kerr, Kerr, and Lincoln 2015b; Franzoni, Scellato, and Stephan 2012). As far as positive selection is concerned, college-educated individuals are migrating three times more than the less-educated in the contemporary world. This ratio drastically varies with economic development at origin. It is slightly greater than one

[6] Cantoni and Yuchtman (2014) show that university training in Roman law played an important role in the establishment of markets during the "Commercial Revolution" in medieval Europe. To establish this, Cantoni and Yuchtman determined the enrollment rates of German students at the universities of Bologna, Paris, Padua, Orléans, Prague, Heidelberg, Cologne, and Erfurt.

in high-income countries, while it reaches 20 in low-income countries (Deuster and Docquier 2019). Such positive selection results from both heterogeneity in incentives and capacity to migrate (Borjas 1987; Chiquiar and Hanson 2005; McKenzie and Rapoport 2007), and immigration policies that favor education and skills.[7]

Migrant selection has also been examined in historical studies, most of them focused on the *Age of Mass Migration* to the United States, a period of unrestricted entry starting in 1850 and ending around 1920.[8] Abramitzky, Boustan, and Eriksson (2012 and 2014) and Spitzer and Zimran (2018) show that selection patterns are consistent with income-maximization models. In the 19th and early 20th centuries, migration to the U.S. was positively selected from some European countries and negatively selected from others. The differences in selection lined up with those in the relative returns to skill across sending countries, or with the easing or tightening of the liquidity constraints (Covarrubias, Lafortune, and Tessada 2015). Using data on servitude contracts from the 17th and 18th centuries, Abramitzky and Braggion (2006) found similar self-selection patterns (on health, physical strength, and literacy) of servants to the American colonies.

Thirdly, we shed light on the mobility patterns at the upper tail of the human capital distribution. Despite the potentially far-reaching implications for international knowledge creation and diffusion (Breschi and Lissoni 2009; Trippl 2013; Miguelez and Moreno 2013; Pierson and Cotgreave 2000), empirical evidence about the drivers and selection of scientists' mobility remains scarce. Existing studies show that, compared to college-educated migrants, scientists and inventors are less sensitive to distance and more sensitive to linguistic proximity, economic conditions, resources dedicated to R&D, and visa-related restrictions (Laudel 2003; Agrawal et al. 2011; Kerr 2008; Fink, Miguelez, and Raffo 2013; Grogger and Hanson 2015). They also show the importance of circular flows which are mostly governed by the existence of scientific collaborations. To the best of our knowledge, none of these studies have focused on the self-selection of scientists. One of the very few studies identifying selection effects among scientists is that of Gibson and McKenzie (2014). Using a survey on the mobility of researchers from the Pacific Islands, they show that current migrants produce substantially more research than similarly skilled return migrants and non-migrants. Hoisl (2007) also shows that mobility is generally found to be positively associated with inventor productivity as proxied, for example, by the education level of the inventor and the use of external sources of knowledge such as university research or scientific literature. Finally, Akcigit, Baslandze, and Stantcheva (2016) find

[7] The structure of migration costs can give rise to many different migration patterns characterized by positive, negative, or intermediate selection. Chiquiar and Hanson (2005), however, observe that Mexican migrants to the United States are drawn from the middle rather than the low end of the Mexican skill distribution, although income inequality is higher in Mexico than in the United States. McKenzie and Rapoport (2007) confirm that Mexican migrants from rural areas mainly come from the middle class of the wealth distribution (those who have both the means and incentives to migrate), and that the intensity of selection decreases with the size of social networks abroad (in line with Beine, Docquier, and Ozden (2011)).

[8] A few studies on intra-European migration support the positive selection hypothesis. Beltrán Tapia and de Miguel Salanova (2017) show that, in the late 19th and early 20th centuries, the literacy level was higher among internal migrants moving to the Spanish capital city than among those who remained in their provinces of origin.

that the international mobility of superstar inventors is influenced by tax policies. A change in one country's top tax rate affects the retention rate of domestic inventors, and has much greater effects on the country's capacity to attract foreign inventors in general and those at the top of the ability distribution in particular.

The remainder of this paper is organized as follows. In Section 2, we present the data sources and define the key concepts used in our analysis. In Section 3, we describe the micro-foundations of our empirical model, present our main findings, and discuss their robustness. In Section 4, we simulate the model to draw its implications for the output of universities per period. The conclusion is in Section 5.

2 Data and Concepts

We collect a large sample of academic scholars (denoted by $i = 1, ..., I$) employed by the universities of Latin Europe over a period that started around the year 1000 CE and lasted until 1800 CE.[9] The year 1800 CE is a convenient date to stop for several reasons. At a broad level, it spelled the end of the Malthusian pre-industrial period. At the university level, it corresponded to profound changes: all French universities were abolished by the Revolution in 1793, and would reappear in a different form later on. In Prussia, the Humboldt reform of 1810 was also a game changer. In this section, we first describe the institutional data sources used to identify academic scholars and their place of work. Secondly, we present the bibliographical data sources used to characterize the lifetime and place of birth of each academic scholar. In the third and fourth parts, we define an index of individual ability or human capital for each scholar, and go into a little more detail with regard to scholars with multiple affiliations. We finally explain how we construct our proxies for institutional notability and quality.

Institutional secondary sources & quality of sampling. − The identification of academic scholars builds mostly on institutions' secondary sources of different types (see Appendix D). Ideally, we aim to cover the universe of scholars involved in university teaching and research before 1800 in Latin Europe. Although this universe is more precisely defined than in other studies of European scholars (e.g. the universe of "famous people" in De la Croix and Licandro (2015), of "creative people" in Serafinelli and Tabellini (2017), or of "notable people" in Gergaud, Laouenan, and Wasmer (2017)), its boundaries remain somewhat flexible. For example, according to biographies of Nicolaus Copernicus, he delivered lectures as a professor of astronomy to numerous students while in Rome. It is unclear whether this teaching took place within the walls of the university of Rome (Sapienza), and how long it lasted. This appears however to be the only time Copernicus taught students. Should we count Copernicus when measuring the notability of the Sapienza? Probably not, as it would overestimate the attractiveness of Rome during this period. Should we include the decision of Copernicus to go to Rome in our study? We did, but it does not matter much as he is only one among thousands of scholars.

[9] Latin Europe means Europe minus the Muslim world and the Byzantine world.

Another dimension of flexibility concerns how we define a university. This seems simple *a priori*. We can rely on Frijhoff (1996) who provides a list of institutions granting doctorate degrees, together with their official foundation date. It is however meaningful to extend this list in two directions. One extension is to include important learning institutions which were not formally universities. One example is the Herborn Academy (*Academia Nassauensis*) which was a Calvinist institution of higher learning in Herborn (Germany) from 1584 to 1817.[10] In addition, another relevant extension is to consider that universities were active before their official creation. For example, the University of Amsterdam was officially founded in 1877, but its roots go back to 1632, when the *Athenaeum Illustre* was founded. For this reason, our period of analysis starts before the official creation date of the first university.

For each university, we first checked whether there is an online historical database of professors. For example, the list of professors at the University of Groningen has already been established. The *Catalogus Professorum Academiae Groninganae* includes all full professors from 1614 onwards (see the website at `http://hoogleraren.ub.rug.nl/`). The website is still under development, but it shows the recent interest of universities themselves in looking at their past in a more systematic way. For those universities without such a database but with books of biographies of their professors, we encoded the contents of these books. For the remaining universities, we checked whether *matricula* (people registered at a given university) and *chartularia* (containing transcriptions of original documents related to the historical events of a university) exist. We have built up a representative sample of professors from this information, and we are continually looking for other national biographies and other databases to complete the information needed. In some cases, the *matriculum* itself is of little use as the status of these people is not recorded (students, professors, etc.), but it follows the chronological succession of rectors, whose names are provided. As rectors were sometimes nominated every six months, their names provide good coverage of the universe of professors there (with some selection bias). Complementary strategies have also been used. For example, for Jesuit universities, there is a biographical dictionary by Sommervogel (1890) listing all Jesuits having published material; as they are classified by place of activity, we can match the professors to the relevant universities. Moreover, for the late Middle Ages and the Renaissance, information can be retrieved from two recent projects, both aimed at collecting biographical and social data on those who graduated from medieval universities: the project “Repertorium Academicum Germanicum - The Graduated Scholars of the Holy Roman Empire between 1250 and 1550” and the project “Studium” for the University of Paris from the 12th century to the Renaissance. Both projects are currently under development.

We grouped universities into three categories, reported in the column “Cov” of Tables 1. The number 3 indicates a comprehensive coverage, i.e. when data collection was based on an

[10]To Frijhoff’s list, we have added the following institutions: the medieval cathedral schools of Chartres and Liège, the school of translators in Toledo, the Majorcan cartographic school, the “Collège Royal” in Paris, the “Jardin Royal des Plantes” in Paris, the Imperial College in Madrid, the Herborn Academy, the Collegium Nobilium of Warsaw, the Gresham college in London, and the Technical University in Braunschweig.

existing website or book whose aim is to list all professors of a given institution. Number 2 indicates broad coverage, often based on the combination of several sources, including books on the history of the university. Number 1 indicates a partial coverage, when the sample of scholars was informed by sources from other universities and general thematic biographies. Notice that the quality of the coverage is not related to the prestige of the university. We have an excellent coverage of the University of Macerata – a small university in Italy, while there is no comprehensive list of professors for the University of Paris. A key requirement of our analysis is to cover almost all scholars with high human capital, and to include a large sample of unknown scholars as well. This requirement is met by encoding the academic scholars included in thematic biographies, such as Taisand (1721) for law, Eloy (1755) for medicine, Junius Institute (2013) for Protestant theology, Herbermann (1913) for important Catholic figures, and Applebaum (2003) for the key actors of the scientific revolution.

Over the whole period 1000-1800, we identify 207 universities and teaching institutions. In the econometric analysis, we eliminate institutions with fewer than 10 scholars or a coverage (total number of professors per year of existence) below 0.05, and thus obtain a working sample of 147 institutions (denoted by $k = 1, ..., K$). Each university k is linked to a geo-referenced location. Accounting for the date of creation of each university, we estimate that these 207 institutions represent a total of 50,317 years of existence. The heterogeneity in the quality and coverage of the institutional data implies that the number of scholars identified varies drastically across universities. A very comprehensive list of scholars can be obtained for the University of Heidelberg which includes 1,186 scholars over 414 years of existence. Note that Heidelberg is not the largest university in our working sample; the data related to the University Bologna allow us to identify 3,290 scholars over the whole period. However, Heidelberg is more representative of an average university than Bologna. Assuming Heidelberg is representative of all institutions, a back-of-the envelope calculation suggests that the order of magnitude of the universe of academic scholars for the medieval and early modern periods is around 144,145 (i.e. 1,186/414 scholars per year × 50,317 years of existence). Observing that scholars taught in 1.11 universities on average, the universe has about 129,431 unique persons.

So far, our bibliographical searches have allowed us to identify 33,726 academic scholars. These include very well-known professors as well as obscure scholars. We thus estimate that our current sample covers around 26.1% of the universe (i.e. 33,726÷129,431). This coverage is very likely to be higher for renowned scholars, as they are more likely to appear in the sources consulted, than for obscure scholars. Having obscure scholars in the sample is important to identify the characteristics of the famous ones – those who are more likely to play the academic market game. Including many obscure scholars in the analysis is thus a strength of our analysis.

Biographical individual data. – We match each scholar's name with bio- and bibliographical dictionaries to identify their place of birth and, later, their quality. We exclude the small number of persons born outside a rectangle encompassing Europe, North Africa, and the Middle East (defined by latitudes $\in [28, 66]$ and longitudes $\in [-22, 51]$) because those would be outliers when computing distances. We also search online for Wikipedia and Worldcat pages to generate the ex post indicators of human capital (see below).

One word about the quality of the bibliographical data. In many cases it is quite high, as the secondary sources used – biographical dictionaries and university sources – were often compiled from archive materials. We should however warn the reader that for the earlier periods, we have chosen to adopt some approximations. A good example is the oculist *Benevenutus Hierosolymitanus*, also called Benevenutus of Jerusalem. His life is totally unknown to historians, but his book, *Ars probatissima oculorum*, was immensely popular and influential – having been translated into four languages already in the medieval period. From other writings citing his work, historians infer he lived between 1100 and 1290. Assigning Jerusalem as his place of birth is disputed, but seems the likeliest option, given the knowledge of Middle Eastern cultures displayed in his writings (remember that Jerusalem was for some time a Latin kingdom (1099–1187)). He was also obviously acquainted with the medical school of Salerno, and he likely taught there (being called the physician from Salerno in one manuscript, the *Besançon Manuscript*). The most intriguing part concerns his relation with Montpellier, another famous medical school. In 1921, the Faculty of Medicine in Montpellier placed a marble slab in its entrance hall listing him among its early faculty members. There are some arguments to link Benevenutus of Jerusalem to Montpellier, but there remains a "considerable disparity between the fragility of the documentary basis for the Montpellier inscription and the robustness of the stone on which it was engraved" (Kedar 1995).

Each individual at university k is characterized by at most five dates: year of birth, year of death, first year of observation at university k, last year of observation at university k, and approximate date of activity at university k (this corresponds to a period or date that is sometimes denoted by "fl.", from the Latin verb *floruit* "s/he flourished").[11] From these dates, we define two dates, t_i^b and t_i^f, which hypothetically bound the active life of each scholar. These dates are computed as follows:

$$\begin{aligned} t_i^b = \min \Big\{ &\text{Year of Birth} + 30, \text{Year of Death}, \min_k[\text{first year of obs. at univ. } k], \\ &\min_k[\text{last year of obs. at univ. } k], \min_k[\text{approximate date at univ. } k] \Big\}. \end{aligned} \tag{1}$$

$$\begin{aligned} t_i^f = \max \Big\{ &\min\left\{\text{Year of Birth} + 50, \text{Year of Death}\right\}, \max_k[\text{first year of obs. at univ. } k], \\ &\max_k[\text{last year of obs. at univ. } k], \max_k[\text{approximate date at univ. } k] \Big\}. \end{aligned} \tag{2}$$

We divide the whole period into eight sub-periods, denoted by

$$\tau \in \{0, 1, 2, 3, 4, 5, 6, 7\},$$

[11]The scholars for which we have no dates cannot be incorporated into the analysis.

corresponding to major historical events: from the urban revolution to the first universities (1000–1199), from the official foundation of Paris and Oxford in 1200 to the Black Death (1200–1347), from the Black Death to the invention of the movable-type printing press (1348–1449), from the printing press to the rise of Protestantism (1450–1526),[12] from Protestantism to the beginning of the Thirty Years' War (1527–1617),[13] from the Thirty Years' War to the revocation of the Edict of Nantes (1618–1684), from this revocation to the rise of Enlightened universities (1685–1733),[14] and from Enlightened universities to 1800 (1734–1800).

We assign each scholar to a period τ based on t_i^b. The beginning date t_i^b should be seen as a time when the individual can make location choices. The final date t_i^f will be used to map the human capital achieved by a scholar to her universities. The period in which this end date falls determines the period for which we impute the quality of the scholar to their university.

Universities' scholars were almost always male, but we found a few females: Trotula de Ruggiero (11th century) and a few others in Salerno, Maddalena Bonsignori, Clotilde Tambroni, Clotilde Zamboni, Bettina and Novella Calderini, and a few others in Bologna, Beatriz Galindo in Salamanca, Ekaterina Romanovna in Moscow, and Dorothea Christiane Erxleben in Halle. Female scholars were a rare exception though. Novellà Calderini, for example, allegedly replaced her father repeatedly, teaching at Bologna veiled so that her beauty would not distract the students, according to the Italian Encyclopedia *Treccani*.

Table 1 shows the number of identified scholars per period, with some of their characteristics. We also report the number of universities per period, which increases steadily except from periods 4 to 6, especially when French Protestant "academies" had to close (Bourchenin 1882). On average, institutional data and bibliographical dictionaries allow us to identify the birthplace of 71.5% of university professors. Hence, we can compute the cost distance d_{ik} associated with each possible scholar-university dyad. Such a cost is defined as $d_{ik} = \ln(\text{cost}^{\min} + \text{cost}_{ij})$ where cost_{ij} is computed using Özak (2010, 2018)'s human mobility index and $\text{cost}^{\min}$ is the minimum cost incurred when having a position in one's own place of birth. We assume it is equivalent to the cost of walking within the old city of Rome between the Vatican City and the Colosseum (3.5 km).

In addition, 23.1% of our identified scholars have a Wikipedia page, and 36.2% have at least one recorded publication in Worldcat. Overall, these shares increase from periods 0-1-2-3 (the Middle Ages) to periods 4-5-6-7 (early modern period). The least well documented period is 1348–1449, when we find many names of professors with no publications, either because they did not publish a lot, were never printed, or their publications did not survive.

Figure 1 shows the university-scholar maps for all sub-periods. Red dots correspond to uni-

[12] 1527 corresponds to the foundation of the University of Marburg, the oldest Protestant university in the world.

[13] This war was of major importance for Germanic universities and the life expectancy of their scholars, as shown in Stelter, de la Croix, and Myrskylä (2021).

[14] In 1734, the University of Göttingen was founded to propagate the ideas of the European Enlightenment.

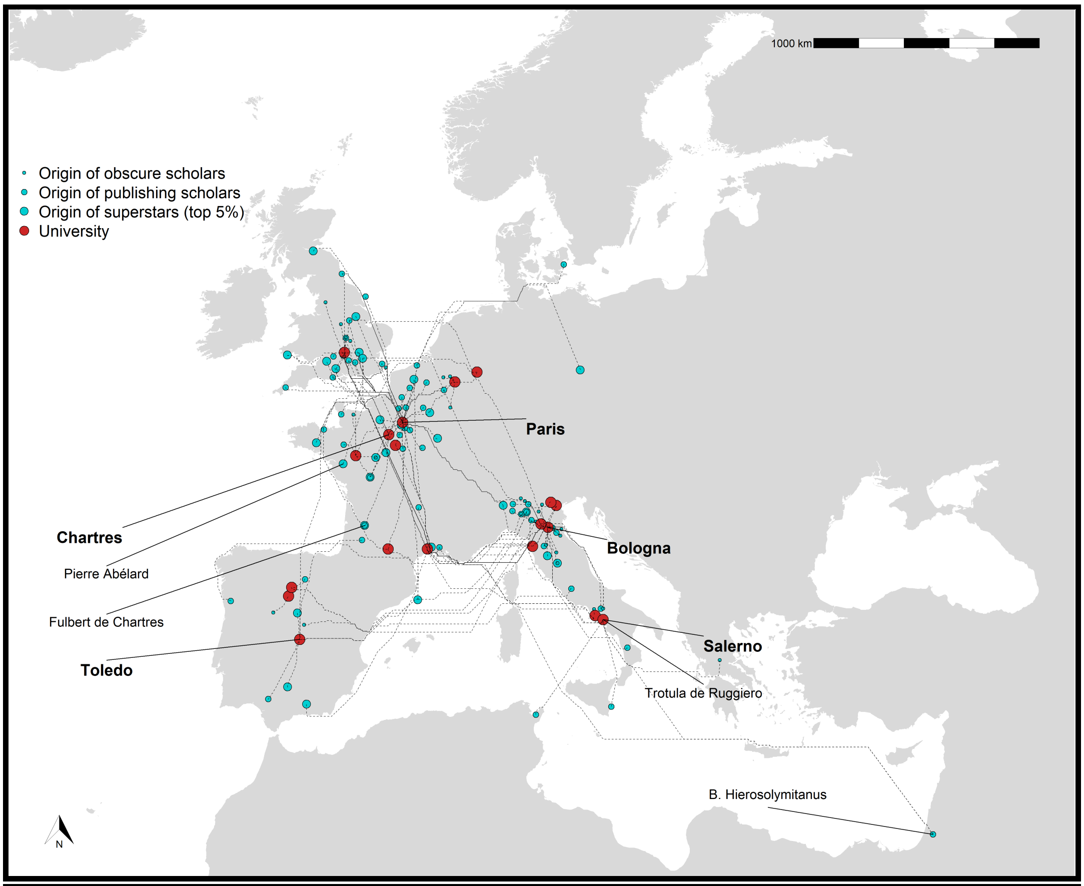
Periods τ		Nb. obs	Nb. univ	Birthplace (%)	Wikipedia (%)	Worldcat(%)
0	1000-1199	268	17	73.1	48.1	47.0
1	1200-1347	1,766	32	62.7	19.9	19.1
2	1348-1449	3,734	49	71.5	10.4	10.0
3	1450-1526	4,838	74	69.5	13.4	17.9
4	1527-1617	7,104	141	75.1	24.4	38.9
5	1618-1685	5,698	137	70.6	24.5	43.6
6	1686-1733	4,294	135	68.1	24.5	46.2
7	1734-1800	6,024	148	74.5	34.6	54.4
Total		33,726	180	71.5	23.1	36.2

Table 1: Summary statistics for professors by sub-period

versities. The top universities are labelled in bold. Blue dots represent scholars' birthplaces and again we have labelled some prominent names. The dashed lines link academic scholars to the university for which they taught. They represent the optimal (i.e. travel-time minimizing) route.

As the first two maps (1000–1347) show, universities emerged in the territory of the late Western Roman Empire. Paris clearly attracted scholars from all over Europe, from Portugal to Scotland and the south of Italy. The density of universities in Italy was already impressive. The period 1348–1449 saw a decline in the number of observations in France, probably due to the Hundred Years' War, combined with the Black Death. West German universities started to play a role, while Italy was very active. We can also see Greek scholars such as John Argyropoulos fleeing the expected fall of the Byzantine Empire (from Harris (1995)). The next period (1450–1526) has the same characteristics, but with more observations in Spain, Scotland, and southern Germany. The number of observations over the period 1527–1617 is high, with good coverage from Portugal to Poland: the portfolio of universities is expanding. The period 1618–1685 saw the development of Nordic universities, and a decline in movement in the south of Europe. A similar trend is observed for the period 1686–1733. The last period 1734–1800 is particularly rich in Germany, and universities expanded to the East. From Iceland comes Grímur Jónsson Thorkelin, who was professor of antiquities at Copenhagen University and is known for the first full translation of the poem Beowulf. On the whole, what can be seen on these eight maps corresponds closely to changes in economic primacy over time in Europe (Kindleberger 1996). More descriptive statistics (including barycenters) can be found in Appendices A, B and C.[15]

[15] We include a breakdown of scholars by broad fields of knowledge. We were surprised to see "theology" decline from 22% to 10.8% between period 0 and period 3 (The Renaissance) and surge again at the occasion of the Reformation, peaking at 21.5% during period 6. It is interesting to contrast this result with the idea that the Reformation led to a secularization of the society. This secularization process is shown in Cantoni, Dittmar, and Yuchtman (2018) through the reallocation of students across fields in Germany (measured by degrees granted and first jobs). Such a reallocation did not seem to be matched by a similar process at the level of the teachers, or might be compensated by more theology in Catholic lands, under the lead of the Jesuits.

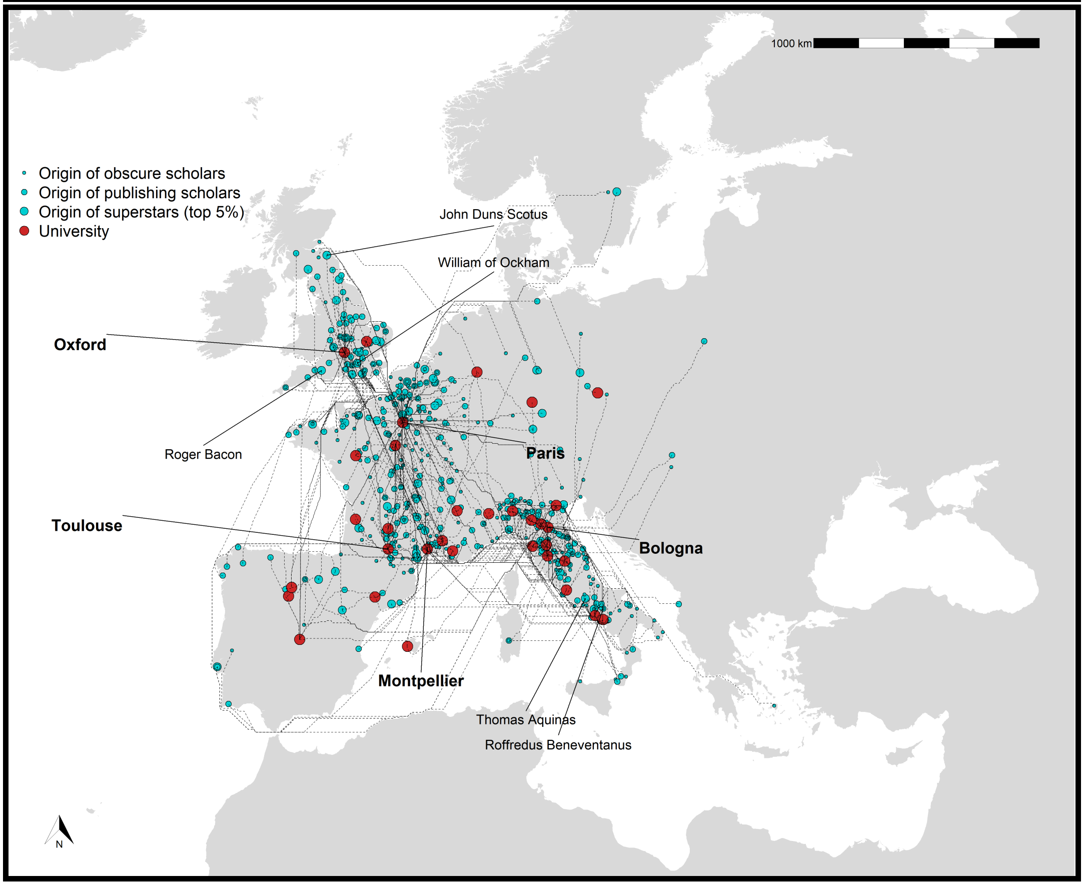


Period 1000-1199

Places of top universities in bold

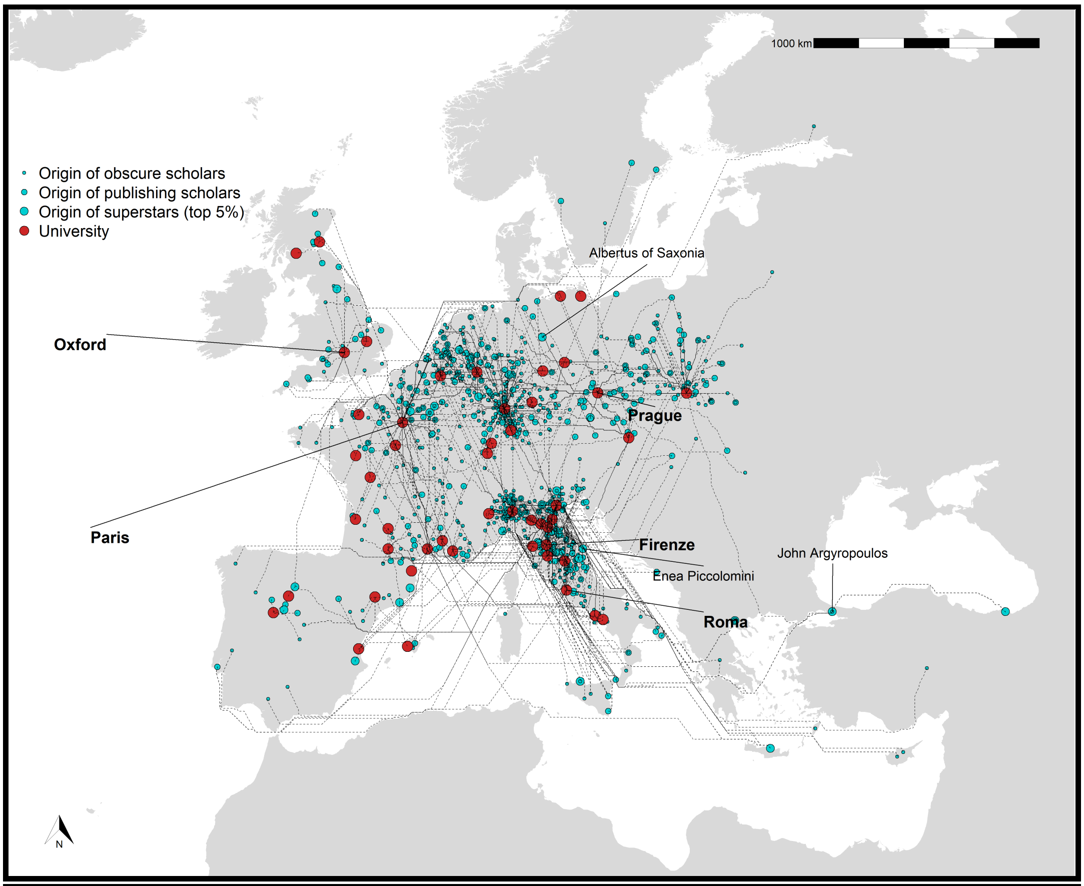
Includes names mentioned in the text

Dashed lines: cost-minimizing path

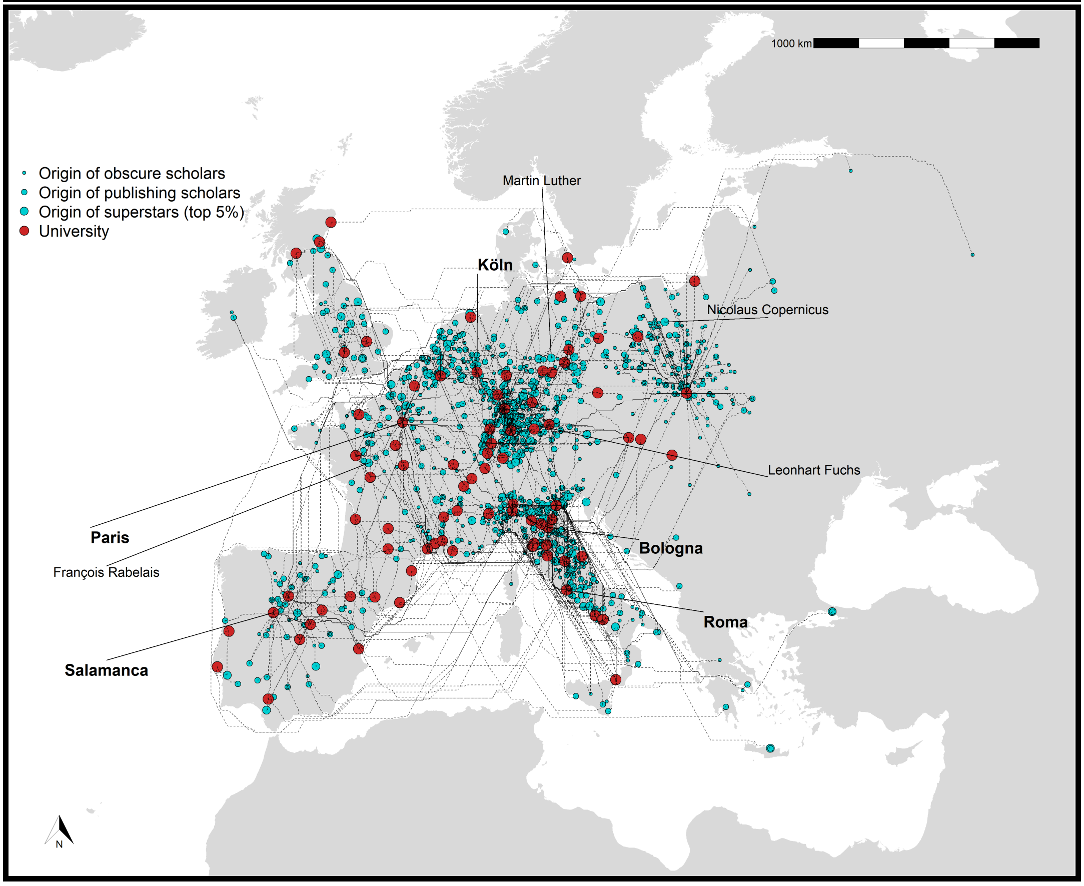


Period 1200-1347

Figure 1: Maps of scholar-university dyads by period (1/4)

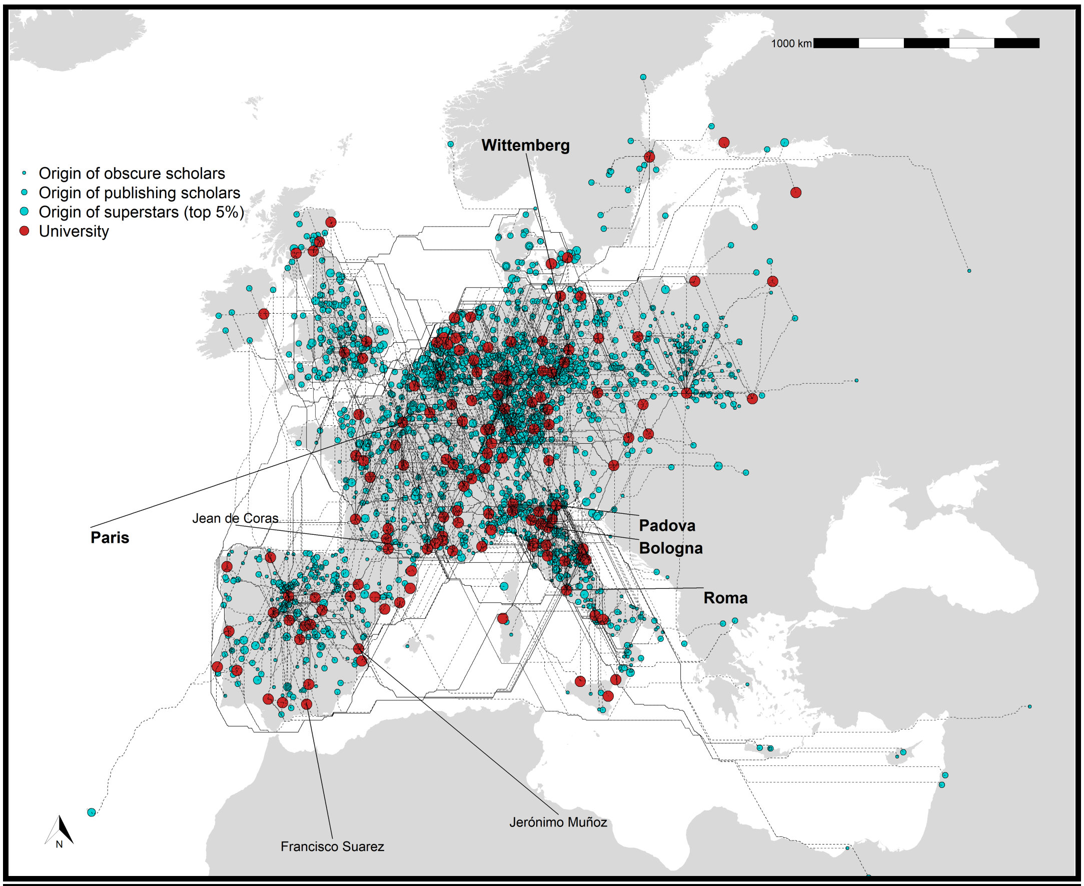


Period
1348-1449

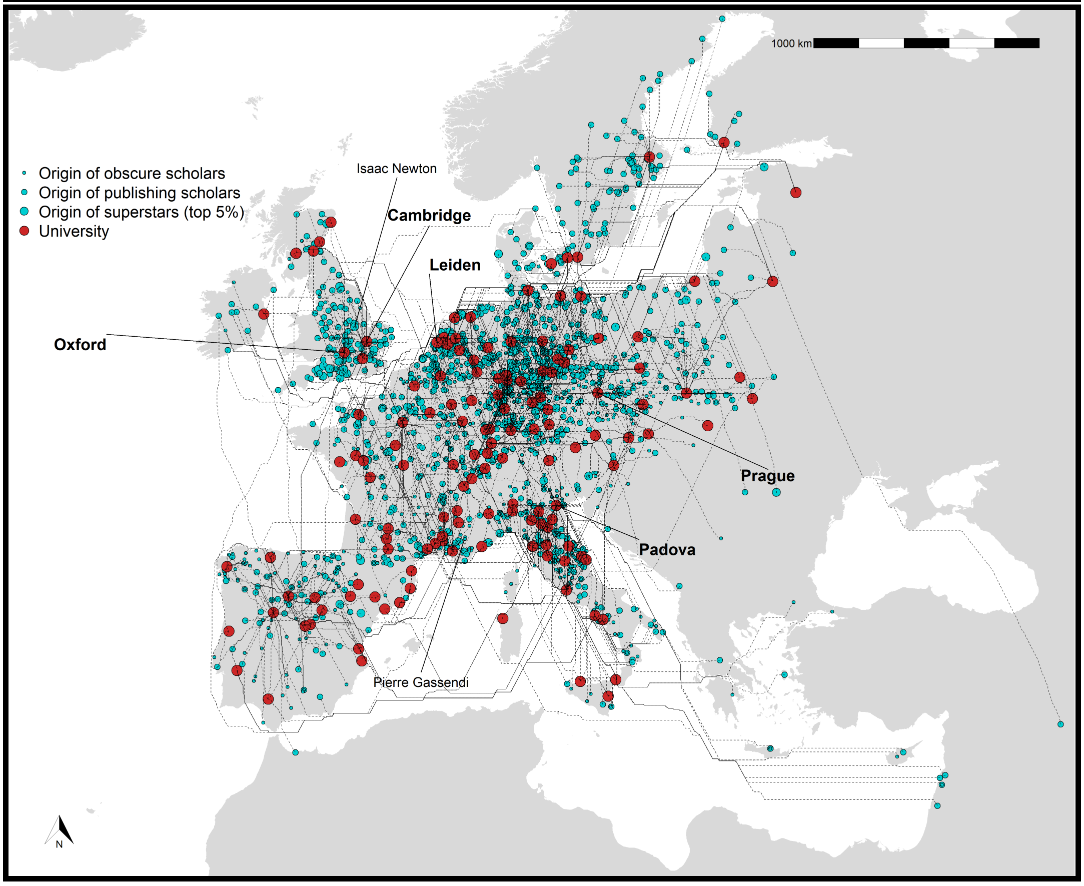


Period
1450-1526

Maps of scholar-university dyads by period (2/4)

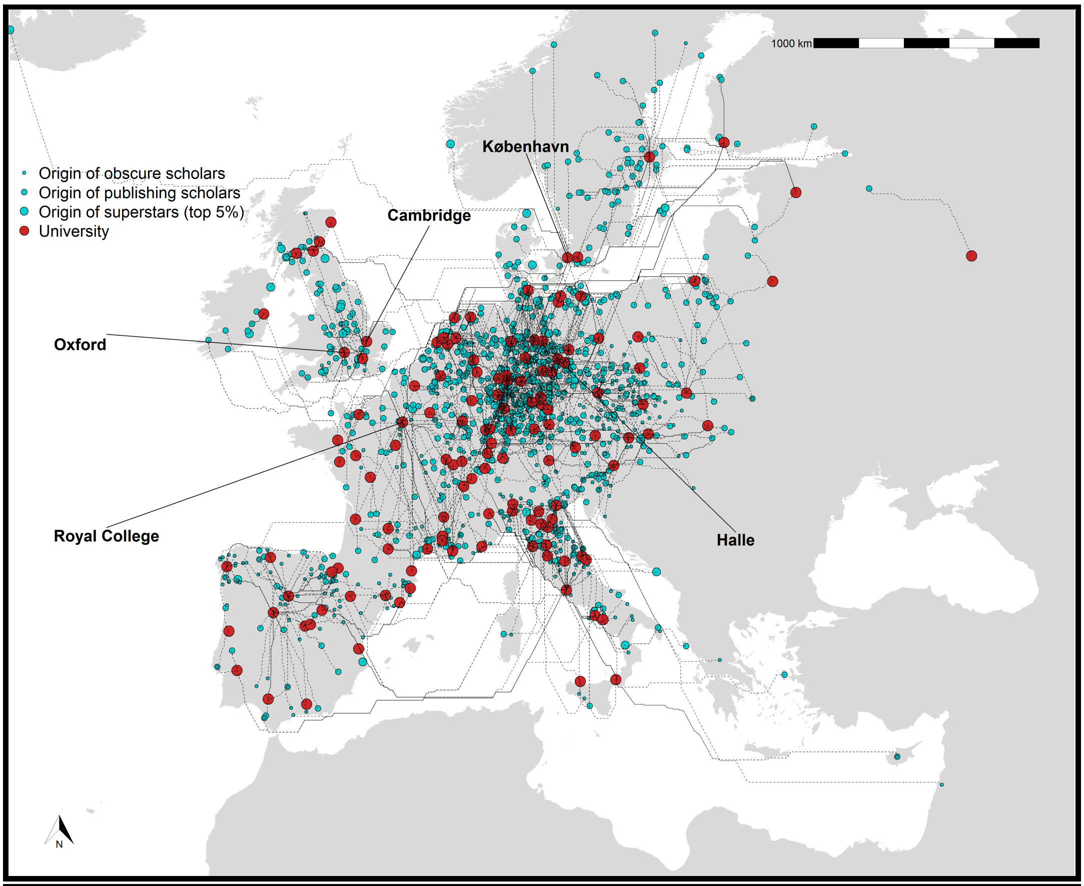


Period
1527-1617

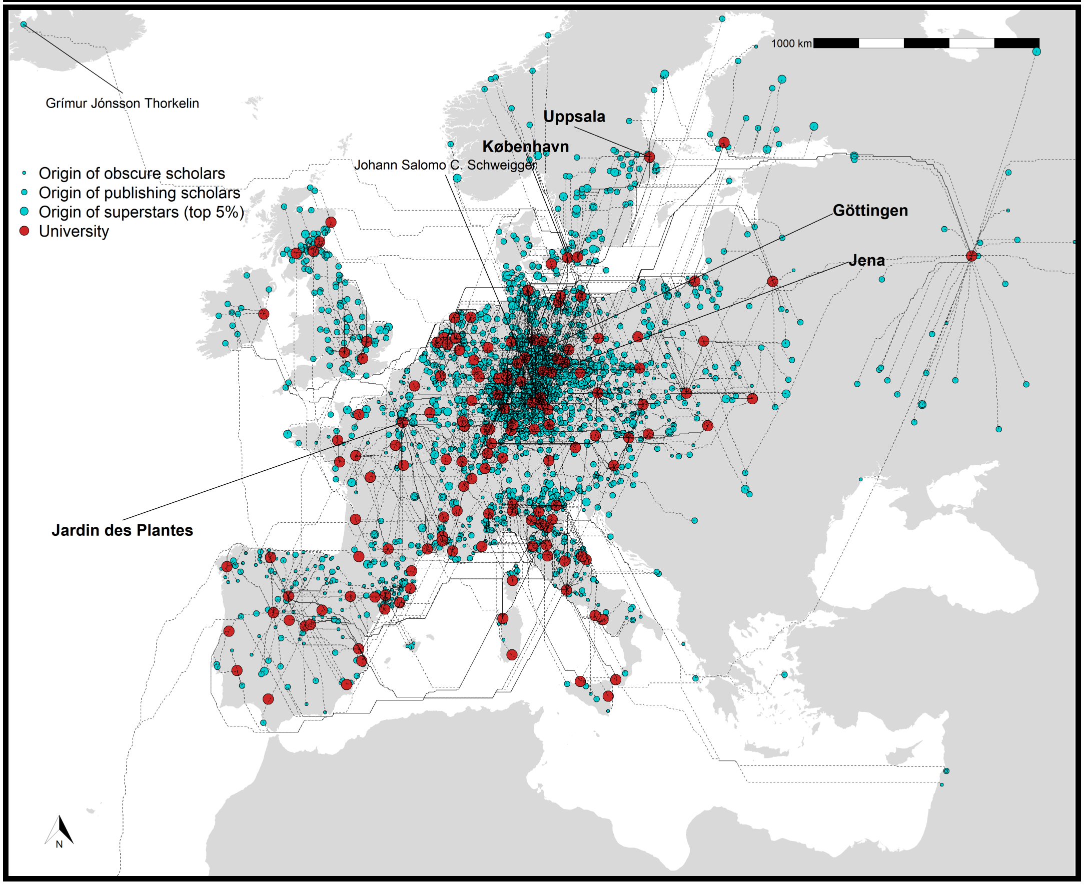


Period
1618-1685

Maps of scholar-university dyads by period (3/4)



Period
1686-1733



Period
1734-1800

Maps of scholar-university dyads by period (4/4)

Using bibliographical data, we define two key concepts that characterize the notability of academic scholars and institutions, and can potentially influence location choices.

Scholars' human capital. – Firstly, we construct an index of *ability or human capital of scholar* i, denoted by q_i. Our index proxies individual notability as seen today in contemporary sources, Worldcat and Wikipedia. Worldcat provides a comprehensive measure of scientific output and citations, as books about the person are included in the measure. Wikipedia completes this measure by putting more weight on the mission of academics called, on today's terms, "service to society" (e.g. becoming an ambassador or a pope, or being canonized a saint). For those who have no Wikipedia and/or Worldcat pages, we have to make two normalization assumptions. We assume first that having no Wikipedia page or a very short Wikipedia page of 60 characters is the same in terms of human capital (the shortest Wikipedia page has 67 characters). Second, we assume that having one publication in one language held by one library worldwide is the same as having no publication at all. To combine the information provided by Worlcat and Wikipedia into one measure, we compute the first principal component of five indicators: (i) the log of the number of characters of the longest Wikipedia page across all languages[16] (ranging from a minimum of 60 to 259,435), (ii) the log of the number of languages in which a Wikipedia page exists (ranging from a minimum of 1 to 220), (iii) the log of the number of works (by or about) in Worldcat (ranging from a minimum of 1 to 74,897), (iv) the log of the number of publication languages in Worldcat (ranging from a minimum of 1 to 52), and (v) the log of the number of library holdings in Worldcat (ranging from a minimum of 1 to 1,083,722).

The results of this analysis are presented in col. (1) of Table 2. The first principal component explains 3/4 of the total variations in the five indicators. The usual heuristic approaches to determine how many principal components one should keep to represent high dimensional data in lower dimensions indicate that one is enough in our case. We finally subtract its minimum value from the first principal component in such a way that a person with no Wikipedia page and no Worldcat entry will have a human capital of zero ($q_i = 0$).

One could argue that a measure of human capital should be based on the works published while the author was still alive. What was published after the death of the person might reflect how the author gained popularity *post-mortem*, which might not be relevant to determining their market value when they were active. It is not possible to implement this because many first editions have not survived. For example, there is no doubt that Pierre Abélard (1079-1142) was a philosopher of great renown during his life.[17] All his written output available in libraries today, from philosophical works to love letters, was published after 1600, and, in many cases, in the last 30 years (see `https://www.worldcat.org/identities/lccn-n79142562/`).

[16]A correction for different languages length was performed, using the translations of the Gospel according to Saint Mark.

[17]Pierre Abélard is also known to the general public for his love affair and correspondence with his pupil Héloïse.

	(1) Benchmark	(2) No Wikipedia	(3) Works by or about
Nb. characters of Wikipedia page	0.429	-	0.395
Nb. languages Wikipedia	0.395	-	0.381
Nb. works in Worldcat	0.471	0.584	-
Nb. languages in Worldcat	0.460	0.562	0.429
Nb. library holdings in Worldcat	0.476	0.585	0.386
Nb. publications by in Worldcat	-	-	0.420
Nb. publications about in Worldcat	-	-	0.435
Nb. Eigenvalues > 1	1	1	1
% variance explained by 1st PC	79.4%	94.0%	76.7%
S.E.	1.993	1.679	2.145
Corr. with (1)	1.000	0.964	0.992
Corr. with (2)	-	1.000	0.945
Corr. with (3)	-	-	1.000

Table 2: First principal component of scholars' human capital

Our measure of q_i is very robust to changes in assumptions. Disregarding Wikipedia leads to col. (2) of Table 2. The correlation between the ability indices computed with and without Wikipedia equals 0.96. In col. (3), we separate the publications by and the publications about the person, and replace the number of works aggregating both types by these two indicators. There is little gain in doing this, and the new measure is correlated with the benchmark with a coefficient of 0.99.

The most famous scholars according to our measure are presented in Table 3 by period. The scholar with the all-time highest human capital is Martin Luther. He was not a scientist like Galileo Galilei, Isaac Newton, or Carl Linnaeus, but it is fair to recognize that he profoundly affected the European sphere. In the list of Table 3, there are some scholars who only have a weak link to a university, and are thus not used to compute the notability of the university. For example, Leonardo da Vinci spent some time at the Studium (university) in Florence to make anatomic dissections; Baruch Spinoza never taught at a university, but interacted with people at the University of Leiden; the same holds for Montaigne and his links with Bordeaux. There are also some who were actual teachers but are better known, at least nowadays, for non-scholarly reasons: François Rabelais, known for his novels, was also in fact a physician who taught at Montpellier; Enea SB Piccolomini (Pope Pius II) or Friedrich von Schiller (German poet) also fall in this category. Their celebrity, even if not strictly academic, was taken into account to compute the notability of the university.

In the same table, we also report the median value of q_i from the set of positive q_i (those with either a Wikipedia or Worldcat reference). It is surprising that there is no visible trend for this

q_i over time, which implies that more recent scholars did not produce more than older ones. It may be more likely that the writings of medieval scholars were lost compared to those of scholars active in the early modern period, yet this loss is compensated for by the accumulation of citations and new editions over time. Let us also note that the particularly high median q_i for the first period, which probably reflects a selection phenomenon, and the low median after the Black Death.

Scholars with multiple affiliations (repeat movers). It is worth noting that our database includes some scholars with multiple career spells or affiliations. Over the whole period, 10.8% of our scholars are linked to more than one university, and the average number of affiliations per scholar equals 1.15.[18] With a record number of 7 recorded affiliations, Jean de Coras and Francisco Suarez are extreme examples of this feature. Jean de Coras (1513-1572) was a French jurist who taught at Padua, Toulouse, Ferrara, Valence, but also, according to Taisand (1721) at Orléans, Paris, and Angers (but we do not even know in which order).[19] Francisco Suarez (1548-1617) was a Spanish Jesuit philosopher and theologian who taught at Avila, Valladolid, Alcala, Salamanca, and Coimbra according to Herbermann (1913), but also at Paris and Rome according to Sommervogel (1890).

We refer to multi-affiliation scholars as repeat movers, and to those who have been only employed by a single university, as one-time movers. It is difficult to make any statement on the reasons for multiple moves. However, there is clear evidence that repeat movers are more likely to belong to the top of the distribution of human capital. Repeat movers are performing better than others at both the extensive and intensive margins. On average, 77.3% of repeat movers have at least one recorded publication, as opposed to 41.7% for one time movers.[20] Focusing on scholars with at least one publication, the average q of repeat movers (4.055) is 25% greater than that of one time movers (3.254). Combining both margins and keeping in mind that the minimal ability level is normalized to zero, the average ability index in the total population of repeat movers (3.167) is 2.3 times greater than the average ability index in the total population of one time movers (1.396). The shares of repeat movers in the population in periods 0 to 7 are equal to 19%, 15%, 8%, 9%, 12%, 11%, 11% and 10%, respectively. The greatest shares, observed in the first two periods, are likely due to a lower coverage of the population of obscure scholars. Although heterogeneity in coverage can skew the comparisons between periods, our data may suggest that the fraction of repeat movers decreased after the Black Death, and increased after the rise of Protestantism. As for the heterogeneity by place of birth, the share of repeat movers varies from 6% in Denmark to 17% in the Netherlands, and 19% in Great-Britain.

[18]If a scholar left a position and came back to the same institution after a while, we consider it as only one affiliation.

[19]Jean the Coras might be known to the international audience as he instructed the famous trial of Martin Guerre. He wrote its best-known record, which was the basis for the movie *The Return of Martin Guerre* with Gérard Depardieu, which was nominated for Best Foreign Language Film by the U.S. National Board of Review of Motion Pictures in 1983.

[20]Those shares are computed on the sample of scholars for whom the birth place in known.

Period τ	**1000-1199**	**1200-1347**	**1348-1449**	**1450-1535**
Med$[q_i\|_{q_i>0}]$	4.45	3.13	2.13	2.72
1	Pierre Ablard (10.1)	Giovanni Boccaccio (11.9)	Nicolas de Cusa (10)	Martin Luther (13.1)
2	Thomas Becket (10)	Thomas Aquinas (11.7)	Johannes Hus (9.7)	Leonardo da Vinci (12.6)
3	Domingo Guzman (9)	Ramon Llull (10.3)	Enea Silvio Piccolomini (9.6)	Desiderius Erasmus (11.8)
4	Rolando Bandinelli (8.5)	John Duns Scotus (10)	Lorenzo Valla (9.2)	Franois Rabelais (11.1)
5	Hugues de Saint-Victor (8.4)	Eckhart von Hochheim (9.9)	Poggio Bracciolini (9.1)	Nicolaus Copernicus (11)
6	Jean de Salisbury (8.4)	Giovianni da Fidanza (9.8)	Georgius Gemistus Pletho (9.1)	Franois Villon (10.3)
7	NA Gratian (8.2)	William of Ockham (9.6)	Francesco della Rovere (8.9)	Ulrich Zwingli (10)
8	Pietro Lombardo (8.2)	Fernando Martins de Bulhoes (9.5)	Jean Gerson (8.9)	Theophrastus von Hohenheim (9.9)
9	Alain de Lille (8)	Marsilio de Padova (9)	Leonardo Bruni (8.9)	Philipp Melanchthon (9.9)
10	Gerardus Cremonensis (7.9)	Roger Bacon (8.8)	Vincent Ferrer (8.7)	Giovanni Pico della Mirandola (9.5)
Period τ	**1536-1617**	**1618-1684**	**1685-1739**	**1740-1800**
Med$[q_i\|_{q_i>0}]$	2.99	2.72	2.74	2.95
1	Galileo Galilei (11.9)	Isaac Newton (11.8)	George Berkeley (10.3)	Friedrich von Schiller (12.6)
2	Francis Bacon (11.3)	Baruch Spinoza (11.4)	Gianbattista Vico (9.7)	Immanuel Kant (12.6)
3	Michel de Montaigne (11.3)	Baltasar Gracin (10)	Ludvig Holberg (9.5)	David Hume (11.3)
4	Johannes Calvin (11.1)	Attanasio Kircher (9.7)	Christian Wolff (9)	Adam Smith (11.1)
5	Johan Kepler (10.7)	Pierre Bayle (9.3)	Hermann Boerhaave (8.9)	Carl Linnaeus (10.9)
6	Torquato Tasso (10.7)	Samuel Pufendorf (9.2)	Johann I Bernoulli (8.7)	Johann Gottlieb Fichte (10.7)
7	Giordano Bruno (10.7)	Thomas Browne (9)	Ruggero Giuseppe Boscovich (8.7)	Christoph Martin Wieland (9.9)
8	Tommaso Campanella (10.2)	Robert Hooke (9)	Daniel Bernoulli (8.6)	Leclerc de Buffon (9.9)
9	Justus Lipsius (9.8)	Jakob I. Bernoulli (8.9)	Johann Christoph Gottsched (8.6)	Bernardin de Saint-Pierre (9.8)
10	Franois Xavier (9.7)	Christopher Wren (8.9)	Richard Bentley (8.5)	Jeremy Bentham (9.8)

Table 3: Top 10 scholars by period with their q_i

Institutional notability. – In theory, we can compute a measure of quality for each university using the observed location and ability levels of the scholars identified in our database. However, given that sampling varies from one institution to the other, computations based on the total number of observed scholars are not directly comparable across places. Taking the means or medians of individual human capital would also be biased in favor of the least well covered universities.

Hence, we introduce the concept of *notability of university* k in period τ as a CES combination of the ability or human capital of the top 5 academic scholars having spent time there and for which the date t_i^f falls in this period. This notability index is denoted by $Q_{k\tau}$. To account for the partial presence of multi-destination scholars, we weight the individual ability q_i by $(1/S_i)^{\omega}$ where S_i is the number of universities where scholar i spent time during their career (i.e. the number of career spells), and we define the adjusted ability level as $\overline{q}_i \equiv q_i\,(1/S_i)^{\omega}$. In our descriptive tables and benchmark regressions, we assume $\omega = 1$ (i.e. the ability of each multi-destination scholar is divided by their number of career spells).[21] We then denote by $(\overline{q}_{1k\tau}, \overline{q}_{2k\tau}, \overline{q}_{3k\tau}, \overline{q}_{4k\tau}, \overline{q}_{5k\tau})$ the ability of the top 5 academic scholars of university k in period τ, and we define the notability index as:

$$Q_{k\tau} = \left(\tfrac{1}{5}\, \overline{q}_{1k\tau}^{\frac{\sigma-1}{\sigma}} + \tfrac{1}{5}\, \overline{q}_{2k\tau}^{\frac{\sigma-1}{\sigma}} + \tfrac{1}{5}\, \overline{q}_{3k\tau}^{\frac{\sigma-1}{\sigma}} + \tfrac{1}{5}\, \overline{q}_{4k\tau}^{\frac{\sigma-1}{\sigma}} + \tfrac{1}{5}\, \overline{q}_{5k\tau}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} + (1-\delta) Q_{k(\tau-1)}, \tag{3}$$

where δ is a depreciation rate that generates some persistence of past notability, and σ is the elasticity of substitution between scholars in producing notability. Hence, $Q_{k\tau}$ is a stock variable. In the benchmark tables and regressions, we assume full depreciation over one period ($\delta = 1$); alternative specifications will be considered in the empirical analysis.

We use $Q_{k\tau}$ as a proxy for the attractiveness of the university. When making location decisions, it is unlikely that scholars were able to accurately quantify the quality of each university. However, they were aware of complementarity forces and they observed the highest ability scholars of each university belonging to their choice set. The notability indices of each university are provided in Appendix D (col. 3 to 10) in the Appendix; we report a blank when there is no scholar at that university during one period (e.g. before the year of creation of each institution) and a zero if all the scholars have $q_i = 0$.

Appendix D lists the institutions kept in the analysis. Compared to the full sample shown in Figure 1, we have removed universities with fewer than 10 scholars in total (as we need enough observations to identify university-specific fixed effects), and also those with extremely low coverage, i.e. fewer than 1 scholar per period of 20 years on average.

Our ranking of the top institutions varies across periods. Prior to 1200, the top universities are Bologna, Paris, Chartres' cathedral school, Salerno, and Toledo's school of translators. From 1200 to 1348, the top universities are Paris, Bologna, Montpellier, Oxford and Toulouse. In the period 1348-1449, we have Paris, Rome (Sapienza), Florence (Studium generale), Prague

[21] We will show below that our results are robust to the choice of ω.

and Oxford. Between 1450 and 1526, we have Rome, Paris, Salamanca, Cologne, and Bologna. Between 1527 and 1617, we have Paris, Wittenberg, Rome, Bologna and Zurich. In the period 1618-1684, we have Leiden, Cambridge, Oxford, Prague, and Padua. Between 1685 and 1733, we have Cambridge, Collège Royal, Copenhagen, Oxford, and Halle. After 1733, we have Jena, Göttingen, Jardins des Plantes, Copenhague, and Halle. This ranking contains a few surprises. For example, the University of Cambridge does very well in periods 5-6, contradicting the view that it was "an intellectual desert, in which a solitary man constructed a system of the world" (Manuel (1968) about Isaac Newton in Cambridge).

One can evaluate the relevance of our ranking of universities by comparing it with rankings obtained using different methods. The Casati Law (Italy, 1858) sets rules for accrediting the pre-existing universities into the new Italian University system (Cottini, Ghinetti, and Moriconi 2019). It ranked universities into three categories, A-B-C depending on their quality. We can compare this ranking with our estimate of Q_{k7}. The average of the Q_{k7} for the 9 universities ranked A is 5.66. The average Q_{k7} for the 8 universities ranked B is 2.18. And the two universities ranked C have a similar level of 2.27 (including the university of Macerata for which we harvested about 800 professors).

3 Empirical analysis

We now turn our attention to the empirical analysis of the determinants of location choices. Economists have long recognized that spatial mobility decisions play a key role in the career choices of workers (e.g. Keane and Wolpin 1997; Neal 1999). Two types of models, spatial search and location choice, have been used to link mobility decisions to career choices. Spatial search and matching models formalize job search decisions across geographically segmented labor markets; they shed light on the effect of distance on the efficiency of a job search, on spatial heterogeneity in search frictions, and on the persistence of labor market disparities between regions (e.g. Manning and Petrongolo 2017; Schmutz and Sidibe 2019). The estimation of matching models requires observing a large number of repeat movers with match-specific outcomes such as individual levels of earnings or employer's profit (e.g. Abowd, Kramarz, and Margolis 1999). This approach in unworkable for us, given the absence of data on match specific outcomes. Moreover, even if we had such outcomes, using only about 10% of the sample (the share of repeat movers) would be costly in terms of external validity of the analysis. Location choice models explain how different types of workers self-select into labor market areas by maximizing their current and expected future levels of income (e.g. Borjas 1987; Dahl 2002; Gallin 2004; Grogger and Hanson 2011). The latter framework is particularly relevant when focusing on the role of workers' attributes, and when match-specific outcomes, demand-side factors and local matching frictions are unobservable. Hence, we opt for this type of framework.

In this section, we first explain the microfoundations and specificities of our location choice model (Section 3.1). We then estimate the determinants of location decisions with a standard logit model is Sections 3.2 and 3.3. The standard logit framework raises a number of econometric issues that might generate inconsistent estimates. Firstly and despite the fact that our database include a large number of obscure scholars, renowned scholars are more likely to be recorded and information about place of birth is missing for a relatively large number of obscure scholars. In the benchmark regressions, these unknowns are eliminated from the sample. This raises sample selection issues that we address in Section 3.4. A related problem is due to the presence of scholars with multiple affiliations. Each (i, k) dyad appears as one observation in the database and is assimilated to a career spell. This means that scholars with seven affiliations appear seven times, while those with a single affiliation appear only once. This also induces possible sample biases and raises the question of the relevance to model scholar i's choice at stage s independently from her other career spells s'. These issues are addressed in Section 3.5. Finally, the benchmark specification disregards the potential endogeneity of q_i, arising from the fact that the ability of scholar i is likely to be affected by her academic environment. We address this issue in Section 3.6.

3.1 A microfounded gravity model

We formalize the discrete location-choice problem of academic scholars in medieval and early modern Europe using a Random Utility Model (RUM), which provides the state-of-the-art microfoundations for most recent gravity models of migration. Our RUM leads to an empirical multinomial logit model which is in line with Akcigit, Baslandze, and Stantcheva (2016), who study the international mobility of superstar inventors since 1977. Standard location choice models assume that the demand-side of the market is perfectly elastic. In our context, this means that the demand for academic scholars (or equivalently, the supply of academic positions) adjusts perfectly to supply. Although most universities have a fixed number of chairs, they also offer a set of other positions which are easily adjusted (e.g. the fellows in Oxbridge, the *professores designati* in Copenhagen (Slottved 1978) , the *survivanciers* (designated successor) in Montpellier (Dulieu 1979)). We account for potential demand-side factors by including "competition costs" whose size depends on the attractiveness of universities and cities as well as on the ability and "market value" of academic scholars.

Compared to the standard literature on the determinants of migration, and beyond the fact that we use unique micro-data, our approach has three specificities. Firstly, we use *geo-referenced* location data. Each scholar i is assigned to a geo-referenced place of birth, whereas each university k is linked to a geo-referenced position. Each scholar-university dyad is associated with a cost distance d_{ik}, measured with the human mobility index (see supra). Since the place of residence of academic scholars before moving to university k cannot be observed, the distance between the place of birth and the university may capture the separation from family and friends (i.e. homesickness), the travel distance *per se*, or the costs of obtaining information about remote places. A striking example of the importance of distance is provided by Eloy (1755) and Michaud (1811) about *Septalius* (Lodivico Settala, 1552-1633). Born and living in Milan, he taught medicine at the nearby University of Pavia and received offers from: the King of Spain, the Duke of Bavaria, the Duke of Tuscany, the city of Bologna, and the Senate of Venice, all offers above what any local citizen could have dreamed of receiving. He enjoyed receiving them as tokens of well-deserved honors, but accepted none. He preferred the company of his fourteen children to the luster of these foreign positions. Another clue to the preference for one's place of birth is the following. Among the 12,997 scholars with a known death place, 812 of them went back to their hometown before dying, although they held appointments in other places during their life. Another 2,618 were born, worked and died at the same place.

Secondly, we exploit the *unbalanced panel dimension* of our database as some scholars made multiple/repeated choices. We do not necessarily know the timing of choices, but our database links several universities to some scholars. We assume an academic career is made of a maximum of S spells indexed by s. At each stage of their career, each professor has to select their preferred location from the feasible university choice set. In practice, if scholar i taught at S_i universities, we include S_i dyadic matches in the database. Robustness checks will be conducted in Section 3.5 to assess the role of movers.

Thirdly, our discrete choice model allows for *varying choice sets*. As new universities are created (or abandoned) over time, the choice sets are individual specific depending on the universities that existed during the active life of the scholar. Each university has a founding date t_0^k and an end date t_1^k, which we mostly take from Frijhoff (1996). Sometimes universities – or some schools which would later become universities – existed before this official date. For example, the University of Paris was officially founded in 1200, but colleges and cathedral schools existed before that date. Gerard Pucelle (1117-1184), an Anglo-French scholar in canon law, taught at Paris from 1156 to 1167 (Arabeyre, Halpérin, and Krynen 2007), before becoming the Bishop of Coventry. We should thus lower the initial date t_0^k for the University of Paris to match the first scholar who can be found there. More generally, the most ancient scholars in the database are Adelbold (965-1027), who taught, at the turn of the millennium, at the cathedral school in Liège, and Fulbert de Chartres (970-1029) who taught at the cathedral school in Chartres and at what would become the University of Angers (Rangeard and Lemarchand 1868).[22] This explains why our period of interest starts in the year 1000 CE. As far as individuals are concerned, we use the time interval $[t_i^b, t_i^f]$ defined in (1)–(2). Let us denote by $\tau(i)$ the period to which professor i is assigned, based on the beginning of her career. The portfolio available to individual i is denoted by $K_{\tau(i)}$. We include a university k in the choice set of individual $K_{\tau(i)}$ if $t_0^k < t_i^f$ or $t_1^k > t_i^b$.

The utility that a professor i obtains from locating at university $k \in K_{\tau(i)}$ at the stage $s \in S$ of her career is given by:

$$U_{isk\tau(i)} = V_{ik\tau(i)} + \epsilon_{isk} = \beta \mathbf{x}_{ik\tau(i)} + \epsilon_{isk}, \tag{4}$$

[22]Both Liège and Chartres had cathedral schools which failed to morph into universities, see Jaeger (2013) on those early cathedral schools in Europe.

where $V_{ik\tau(i)} = \beta \mathbf{x}_{ik\tau(i)}$ represents the deterministic component of the indirect utility (net of moving costs), which depends on a vector of observable variables, and ϵ_{isk} is a vector of person-specific random taste shocks representing the unobservable determinants which enter the utility functions and are orthogonal to the deterministic component.

Assuming the random term ϵ_{isk} is independently and identically distributed as Extreme Value Type I (EVT-I), which implies that multiple career choices are independent, we can model the probability that university k represents the utility-maximizing choice for professor i at the stage s of her career as the outcome of a standard multinomial logit model (McFadden 1974):

$$p_{isk\tau(i)} \equiv \text{Prob}\left[U_{isk\tau(i)} = \underset{k' \in K_{\tau(i)}}{\text{Max}}\, U_{isk'\tau(i)}\right] = \frac{\exp(\beta \mathbf{x}_{ik\tau(i)})}{\sum_{k' \in K_{\tau(i)}} \exp(\beta \mathbf{x}_{ik'\tau(i)})}. \tag{5}$$

In this formula, the probability of going to a given place depends on the features of that place (the numerator) compared to the features of all the other places in the portfolio (the denominator). The property of the multinomial logit model is that the relative probability of choosing between two alternative options in $K_{\tau(i)}$ depends on the attractiveness of these two options only, i.e.

$$\ln p_{isk\tau(i)} - \ln p_{isk'\tau(i)} = \beta \mathbf{x}_{ik\tau(i)} - \beta \mathbf{x}_{ik'\tau(i)},$$

and is independent of the presence of other alternatives (IIA: Independence of Irrelevant Alternatives). In addition, the choice probabilities are independent across career spells as long as ϵ_{isk} and $\epsilon_{is'k}$ are assumed to be independently distributed. The latter assumption will be relaxed later.

As in the literature on migration, in which the location choice of migrants conditional on the decision to migrate (Bertoli and Ruyssen 2018) is studied, our estimations are conditional on the choice of having an academic career. As we cannot observe the universe of scholars, including those not choosing to teach at universities, we cannot model the *ex ante* problem of choosing between universities and other activities. Notice that this choice is more complex than a binary choice, as many scholars combined positions at universities with other occupations (such as physician or astronomer to the monarch, bishop or judge). Our estimation thus rests on the independence of irrelevant alternatives property within the choice set $K_{\tau(i)}$, which implies that the relative probability of choosing between two alternative options in $K_{\tau(i)}$ depends exclusively on the attractiveness of these two options. Even if selection into academia would not affect the location choice of individuals having chosen to teach, it might affect our simulations if – for example – the total number of professors depends on the notability of universities. Hence it is fair to acknowledge that our results remain partial equilibrium results.

Estimating the multinomial logit (5) requires specifying the analytical form of the deterministic component of the utility function as a function of observable individual (q_i), institutional ($Q_{k\tau(i)}$), and dyadic characteristics (d_{ik}). In the benchmark model, we consider q_i as independent of her location choice. We also first consider $Q_{k\tau(i)}$ as exogenous, although we adjust it to eliminate the influence of any scholar i on the notability of the university. For each scholar i, we exclude their own ability index from the calculation of the university notability index. We thus make this notability index person-specific, $Q_{ik\tau(i)}$, to mitigate endogeneity concerns. The endogeneity of individual ability and adjusted institutional notability (influenced by the potential spillovers of scholar i on her colleagues) will be treated later.

The deterministic component of the utility function captures the average benefits and the average cost for i of locating at k, and is independent of the career spell s:

$$V_{ik\tau(i)} \equiv B_{ik\tau(i)}(.) - C_{ik\tau(i)}(.). \tag{6}$$

We model the benefits ($B_{ik\tau(i)}$) as an increasing function of the attractiveness of the city where the university is located (proxied by the population density, $P_{k\tau(i)}$, and by the indicator of local democracy from Bosker, Buringh, and Van Zanden (2013), $D_{k\tau(i)}$), as well as of the adjusted notability of the university ($Q_{ik\tau(i)}$), as suggested by anecdotal evidence. For example, Navarro-Brotons (2006) discusses the case of Jeronimo Munoz, who moved from Valencia to Salamanca in 1578. Although Munoz was one of the best paid professors at the University of Valencia, his salary was considerably lower than those paid at universities in Castille. The prestige of the University of Salamanca, and its greater proximity to the seat of royal power, was probably also a factor in Munoz's decision to accept Salamanca's offer. Furthermore, the effect of $Q_{ik\tau(i)}$ can vary with the ability of the professor as, for example, high-ability professors benefit more (or less) from expected interactions with high-ability colleagues (e.g. Stephan and Levin 2001; Kerr, Kerr, and Lincoln 2015a; Kerr, Kerr, and Lincoln 2015b; Kerr et al. 2017). We assume the following specification:

$$B_{ik\tau(i)} = a_0 + a_1 Q_{ik\tau(i)} + a_2 P_{k\tau(i)} + a_3 D_{k\tau(i)} + a_4 q_i Q_{ik\tau(i)} \tag{7}$$

where all coefficients are expected to be non-negative.

We model the cost of locating at university k ($C_{ik\tau(i)}$) as an increasing function of the cost distance from the place of birth (d_{ik}) and of the competition for finding a job at university k in period $\tau(i)$. The competition for finding a job reflects the demand side of the academic market. Again, anecdotal evidence suggests that the recruitment policy of the best universities included efforts to attract international talent. To give two examples, Eloy (1755) reports that Leonhart Fuchs (after whom the plant fuschia was named), a professor at Ingolstadt in 1526, was offered six hundred gold coins by the Duke of Tuscany, Como, to teach at the University of Pisa. Nadal (1861) discusses the case of the University of Valence, which was searching for a renowned legal scholar in 1583. They sent a messenger to convince a famous lawyer in Grenoble, Jean-Antoine de Lescure, to join the university. The latter reported that he would be willing to come for a salary of 1,500 pounds, provided his moving and house rental costs were covered by the university. They finally agreed on 1,200 pounds plus the house, partly paid by four merchants of the city. Later on, his colleague François Josserand became jealous of Lescure's treatment, threatened to go elsewhere, and obtained a pay rise.

We reasonably assume that the "competition cost" incurred by a professor increases with the attractiveness of the city ($P_{k\tau(i)}$ and $D_{k\tau(i)}$), as well as with the adjusted notability of the university ($Q_{ik\tau(i)}$). However, we also allow the latter "competition cost" to be negatively affected by the individual level of ability, as high-ability professors have a higher market value and receive more generous offers from top universities. In line with the literature on self-selection in migration (e.g. Grogger and Hanson 2011; Beine, Docquier, and Ozden 2011; Kerr et al. 2017; Kerr et al. 2016), we allow the cost of distance to be negatively affected by the individual level of ability. We assume the following specification:

$$C_{ik\tau(i)} = b_0 + b_1 Q_{ik\tau(i)} + b_2 P_{k\tau(i)} + b_3 D_{k\tau(i)} - b_4 q_i Q_{ik\tau(i)} - b_5 d_{ik} + b_6 d_{ik} q_i \tag{8}$$

where all b's are expected to be non-negative.

Plugging (7) and (8) into (6) gives the expression for the net benefit of an (i, k) employment match. However, in our empirical regressions, we extend the number of generic determinants of location choices ($\mathbf{x}_{ik\tau(i)}$) to account for the imperfect coverage of our database and for unobserved heterogeneity. We add a university fixed effect, γ_k, which captures both the unobserved pull factors associated with university/city k that do not vary across periods and the quality and extent of the sources used for each university. This yields:

$$V_{ik\tau(i)} \equiv \beta \mathbf{x}_{ik\tau(i)} = \beta_0 + \underbrace{\beta_1 Q_{ik\tau(i)} + \beta_2 P_{k\tau(i)} + \beta_3 D_{k\tau(i)}}_{\text{Agglomeration}} + \underbrace{\beta_4 q_i Q_{ik\tau(i)}}_{\text{Sorting}} + \underbrace{\beta_5 d_{ik}}_{\text{Distance}} + \underbrace{\beta_6 d_{ik} q_i}_{\text{Selection}} + \gamma_k \tag{9}$$

where β is a set of parameters that are common to all individuals and that can be estimated. Unlike standard (linear) regression models, the specification of the multinomial logit model depicted in Eq. (5) implies that the individual probability to take a position in a university k depends on the characteristics of all universities and cities. Any change in one of these characteristics impacts the whole system.

Identification. In line with (7) and (8), the constant is given by $\beta_0 \equiv a_0 - b_0$. Coefficients $\beta_1 \equiv a_1 - b_1$, $\beta_2 \equiv a_2 - b_2$ and $\beta_3 \equiv a_3 - b_3$ can be positive or negative and reflect the *agglomeration* (or dispersion) effects resulting from the attractiveness and competition effects. As the university fixed effect captures the mean level of agglomeration/dispersion forces throughout the entire period covered by our sample, our estimation of these coefficients exploits the within (or demeaned) variations over time in the notability of universities and in the attractiveness of cities. Coefficients β_1, β_2 and β_3 are identified by the fact that the number of scholars who decide to take a position in university k decreases when this university becomes worse in terms of quality (e.g., Spanish universities in periods 6 and 7) and when a city loses its communal freedom (e.g., Northern Italian cities after the Renaissance); by contrast, it increases when a city becomes relatively bigger (e.g., Amsterdam in periods 6 and 7).

Just as $V_{ik\tau(i)}$, the other determinants of location choices are dyadic by construction (sorting and selection terms) or in nature (distance), which makes their identification and estimation possible. Coefficient $\beta_4 \equiv a_4 + b_4$ is positive if high-ability scholars tend to agglomerate at better universities (what we refer to as *positive sorting*) due to higher benefits or smaller costs; it is identified by the fact that the best scholars are more likely to take a position in university k when its quality increases. Coefficient $\beta_5 \equiv -b_5$ is the standard *Distance* term capturing the expected negative effect of remoteness; it is identified by the fact that a given university attracts more scholars born in its vicinity than born far away. As for $\beta_6 \equiv b_6$, it is positive if high-ability individuals are more mobile than lower-ability ones (what we refer to as *positive selection*);[23] it is identified by the fact that renowned scholars are less sensitive to distance and are more likely to take a position in a remote university than obscure scholars.

Finally, the multinomial logit expression (5) implies that variables that are not specific to a destination k, directly or through their interaction with individual characteristics, cannot be identified, as they would affect the net benefit of all (i, k) employment matches symmetrically. This explains why our set of regressors in (9) includes neither purely personal characteristics (such as the ability of scholar i *per se*) nor purely temporal phenomena (such as time fixed effects).

3.2 Results from the multinomial logit model

Table 4 contains the results of a standard multinomial logit regressions for the whole period 1000-1800. The estimations are obtained by using the mlogit package of Croissant (2012), which allows for varying choice sets. These regressions characterize the location choices of 23,624 scholars with a mean number of career spells equal to 1.15, which gives a total of 27,145 individual observations. Denoting the number of elements in set S, by $\overline{\overline{S}}$, the mean number of institutions is equal to $E_i\overline{\overline{K_i}} = 100$ (the total number $\overline{\overline{\cup_i K_i}} = 147$), implying that our database includes 2,724,714 possible dyadic matches. We focus here on the sign and significance of the agglomeration, distance, selection, and sorting terms. In all regressions, we control for institution fixed effects. The sizes of cities, $P_{k\tau(i)}$ are obtained from Bairoch, Batou, and Chevre (1988) with the following mapping between periods τ and dates available in Bairoch et al.: 0:1000, 1:1200, 2:1400, 3:1500, 4: average between 1500 and 1600, 5:1600, 6:1700, and 7:1750. The level of local democracy, $D_{k\tau(i)}$, is obtained from Bosker, Buringh, and Van Zanden (2013) who created a binary variable equal to one when cities could organize themselves and claim a kind of self-rule that was often acknowledged by the sovereign in return for taxes or loyalty. The first occurrences of communal self-government were identified in the 11th and 12th centuries in Spain and Italy. They spread across the rest of Europe in the following centuries.

The regression in col. (1) can be seen as a textbook gravity equation, including distance d_{ik} and mass (in the fixed effect γ_k). This standard gravity regression shows that the probability of observing a scholar-university match decreases with the cost distance between the birthplace and the university location. This effect remains strong in all specifications. The coefficient of

[23]Positive selection and sorting can also arise if the utility function (4) is not additively separable.

	(1)	(2)	(3)	(4)	(5)	(6)
Distance:						
d_{ik}	-1.698***	-1.693***	-1.762***	-1.691***	-1.752***	-1.757***
	(0.007)	(0.007)	(0.009)	(0.007)	(0.009)	(0.008)
Agglomeration:						
$Q_{ik\tau(i)}$		0.216***	0.217***	0.180***	0.184***	0.310***
		(0.006)	(0.006)	(0.007)	(0.007)	(0.005)
$P_{k\tau(i)}$		0.164***	0.164***	0.168***	0.168***	-0.006
		(0.023)	(0.023)	(0.022)	(0.023)	(0.007)
$D_{k\tau(i)}$		0.145***	0.147***	0.140***	0.142***	0.217***
		(0.038)	(0.039)	(0.038)	(0.039)	(0.018)
Selection:						
$d_{ik}q_i$			0.033***		0.030***	0.052***
			(0.003)		(0.003)	(0.003)
Sorting:						
$Q_{ik\tau(i)}q_i$				0.017***	0.015***	0.009***
				(0.001)	(0.001)	(0.001)
k FE	yes	yes	yes	yes	yes	no
N. Obs.	27,145	27,145	27,145	27,145	27,145	27,145
Log Likelihood	-59,949	-59,126	-59,052	-59,057	-58,999	-64,402

Table 4: Multinomial logit regressions: results from a standard logit model

distance is above unity, $\hat{b}_5 = -\hat{\beta}_5 > 1$, which is unsurprisingly greater than in the contemporary period. Focusing on the stock of international migrants in 2010, Beine, Docquier, and Ozden (2011) find a coefficient of 0.7 for all migrants and of 0.35 for college-educated migrants. Focusing on current academic researchers, Fink, Miguelez, and Raffo (2013) find a smaller coefficient around 0.2. Agglomeration forces are added in col. (2). Scholars are attracted by the notability of the university, the size of the city, and the level of local democracy. Notice that due to the presence of university fixed effects, the agglomeration effects are identified through the variations in institutional notability, city size, and democracy over time, while the effect of distance is identified through the spatial variation in d_{ik}. In col. (3), we add the interaction between distance and individual human capital $d_{ik}q_i$. This term is positive, which suggests that the most notable professors were more mobile than others (*positive selection*). In col. (4), we interact the individual human capital index with the notability of the university. We find evidence of *positive sorting*: the most notable professors were more likely to settle in more prestigious universities. Putting all regressors together in col. (5) shows that agglomeration, selection, and sorting are all significant. Using the values of the log likelihoods, we can compute some simple LR tests: comparing (2) to (1), we can reject the null hypothesis that there is no agglomeration effect. Similarly, comparing (5) to (2), we reject the absence of selection and sorting. To illustrate, university fixed effects are excluded in col. (6); all results are similar with the exception of the local population effect, which becomes insignificant.

To determine whether the coefficient of distance is stable over time, we also ran a specification with distance interacted with a period dummy. This allows us to test whether the speed of travel improved before 1800. The eight estimated coefficients are: -1.225, -1.301, -1.684, -1.737, -1.698, -1.816, -1.843, and -1.850. The coefficient is thus quite stable over the last 5 periods, but lower during the Middle Ages, and especially during the periods before the Black Death. The other coefficients are unaffected, except the effect of communal freedom, which is reinforced. The unexpected non-decreasing pattern in this coefficient reflects that there was little progress in the quality of roads until the 18th century (Bogart 2011), and little innovation in travel by boat before the invention of steamboats in the 19th century. The lower cost of moving during the Middle Ages may reflect weaker national states, and also the lower density of universities in this period.

The coefficient of the interaction term $Q_{ik\tau(i)}q_i$ captures the fact that high-quality scholars are more sensitive to the reputation of the university when solving the location-decision problem that they face, and/or that higher-quality universities reward scholars' quality more (i.e. higher wages per unit of quality). Wages are unobserved for us. Assuming that wages are proportional to q_i, we may want to include q_i among the determinants of location-specific utility, allowing its coefficient to vary across alternatives. This is standard in the estimation of a multinomial logit model with variables that are individual but not alternative specific. Still under the assumption that wages are proportional to q_i, it would purge the estimated coefficient of $Q_{ik\tau(i)}q_i$ from the confounding effect of differences in wages across universities. Including these choice-specific terms, we obtain $\overline{\overline{\cup_i K_i}} = 147$ more parameters to estimate. The estimated coefficients of these q_is vary from one university to the other, as does the university fixed factor. They also sometimes have a negative value, which is hard to interpret in the context where scholars would be remunerated in proportion to their q_i. In this new specification, the interaction term $Q_{ik\tau(i)}q_i$ is weakened but remains highly significant (0.007 (0.002) instead of 0.015 (0.001)) despite the inclusion of many terms correlated with q_i. The three agglomeration effects are barely affected.

In a non-linear model, the coefficients cannot be interpreted in terms of predicted probability as the effect of a change in a variable depends on the values of all variables in the model. To put it differently, the effect depends on where we evaluate it: the derivatives of the choice probabilities are given by $\frac{\partial p_{isk}}{x_{ik}} = \beta p_{isk}(1 - p_{isk})$, which is largest when $p_{isk} = 0.5$. For this reason, our coefficients β can only be interpreted as the effect of x_{ik} on indirect utility. This will be very clear at the beginning of the next section, where we will simulate the model with and without selection and sorting for a person with a high q_i, and compare with another one with a low q_i. The results in Table 4 also indicate that the effect of positive selection is relatively small: when q_i is around 10 (scholars at the top of the ability distribution), the utility loss due to distance is reduced by just 10%. By contrast, the effect of positive sorting is large: when q_i is around 10, sorting increases the gains from settling in a more prestigious university or in a more attractive city by a factor of 2 to 3. Besides the standard distance term, agglomeration and positive sorting are important forces governing the mobility decisions of academic scholars. Table 2 in Appendix F shows that our estimation results are highly robust to the choice of parameters (δ, σ, ω).

3.3 Heterogeneous effects

The benchmark assumption of a constant university fixed effect across fields of knowledge and across periods (γ_k) is made for simplicity. In practice, the attractiveness of a university varied over time and across fields of study. To solve this problem, we separately re-estimate the multinomial logit (5) after excluding some periods, some fields of study, and some regions of birth. As for the fields, we distinguish between Theology, Law, Medicine, and Science. The field(s) of a scholar are mostly identified through the courses taught. Law includes both canon and civil law. Medicine includes Anatomy, Surgery, and Pharmacy. Sciences include Mathematics, Logic, Physics, Chemistry, Biology, Astronomy, Earth Science, Geography, and Botany. One should be aware that the distinction between these fields is a bit arbitrary, in particular when going back in time. For example, the theologians Thomas Aquinas and Albertus of Saxonia spent time reconciling the Aristotelian view of a finite world with the Christian view of an infinite God. In doing so, they contributed to the development of the mathematical notion of limit (Sergescu 1939). Pierre Gassendi (1592-1655) is known as an astronomer (a crater on the moon was named after him), but was a professor of theology at Aix-en-Provence.

In Table 5, we separately estimate the multinomial logit (a) for five sub-periods (i.e. 1000 to 1526, 1200 to 1617, 1348 to 1685, 1450 to 1733, and 1527 to 1800), (b) after excluding one field at a time (i.e. Theology, Law, Medicine, and Science), and (c) after excluding one region of origin at a time. We successively exclude the Low Countries (Benelux), Germany (as of today), France (as of today), Italy, and the British Isles (currently the United Kingdom and Ireland). Our 14 sub-samples include smaller numbers of observations. In each of these sub-samples, the portfolio of possible universities differs. For example, when we exclude scholars born in Germany, some German universities cannot be included in the estimation as only German scholars worked there during their existence.

The effect of distance is always negative and highly significant. As far as agglomeration forces are concerned, the notability of the university is always positive and significant. The attracting effect of city size is always positive, with the exception of the first period. The effect of communal freedom, which is found to be important in general by Serafinelli and Tabellini (2017) in their study of the migration patterns (from birth to death) of creative people, is found to be significant in most sub-samples. The estimates by region show that Italy is key to identifying this effect, as communal freedom stops being significant when one removes Italian scholars from the sample. Italian cities are unique in that many lost their freedom during the Renaissance. This echoes the study of Buonanno et al. (2019) who show that territories with communal freedom in the Middle Ages display more positive features and attitudes today than

	Distance	Agglomeration			Selec	Sorting	Nb
	d_{ik}	$Q_{ik\tau(i)}$	$P_{ik\tau(i)}$	$D_{ik\tau(i)}$	$d_{ik}q_i$	$Q_{ik\tau(i)}q_i$	of obs.
Benchmark	-1.752***	0.168***	0.168***	0.142*	0.030***	0.015***	27,145
	By sub-period						
1000-1526	-1.661***	0.278***	-0.126***	0.244***	0.049***	-0.003	8,084
1200-1617	-1.712***	0.273***	0.015	0.222***	0.046***	-0.002	13,938
1348-1685	-1.761***	0.153***	0.244***	0.131***	0.036***	0.004**	17,164
1450-1733	-1.804***	0.135***	0.374***	0.410***	0.038***	0.013***	17,599
1527-1800	-1.795***	0.074***	0.280***	0.360***	0.029***	0.033***	18,936
	By field						
W/o Theology	-1.748***	0.195***	0.240***	0.172***	0.031***	0.013***	21,769
W/o Law	-1.754***	0.174***	0.145***	0.187***	0.021***	0.019***	18,778
W/o Medicine	-1.735***	0.191***	0.181***	0.129***	0.027***	0.014***	22,747
W/o Sciences	-1.773***	0.195***	0.164***	0.138***	0.035***	0.012***	24,153
	By region						
W/o Benelux	-1.741***	0.179***	0.176***	0.131***	0.031***	0.015***	25,810
W/o Germany	-1.714***	0.209***	0.095***	0.153***	0.028***	0.011***	20,836
W/o France	-1.698***	0.184***	0.170***	0.162***	0.020***	0.018***	23,764
W/o Italy	-1.739***	0.164***	0.257***	0.079	0.023***	0.020***	17,805
W/o UK/Irl	-1.752***	0.179***	0.160***	0.132***	0.039***	0.013***	25,906

Table 5: Multinomial logit regressions: heterogeneous effects

territories without such freedom. As for positive selection, the effect is positive and significant in all cases, more prevalent for theology and sciences. Finally, the sorting term is positive and significant in all cases as well. Despite smaller numbers of observations, our results are fairly robust across sub-samples.

The results by period can be used to consider the effect of Protestantism on the academic market. The period 1000-1526 ends with the creation of the first Protestant university, Marburg. The period 1527-1800 covers a divided world, where many scholars had to change religions if they wanted to keep their positions, while others decided to migrate to a region where their religion was accepted. Others converted voluntarily and this changed their approach to science.[24]

In the period before the Reformation, the agglomeration force attracting all scholars to the most notable universities is very strong (coefficient of $Q_{ik\tau(i)}$ around 0.278). Positive sorting seems negligible then. In the post-Reformation period, it is the opposite. The agglomeration force weakens (the coefficient is about 0.074), but sorting is strong, indicating that the ability of top

[24] An example is Nicolas Steno (from Table 3). Born to a Lutheran family and known for his groundbreaking contributions to geology, he converted to Catholicism and moved away from the natural sciences to embrace theology.

universities to attract professors became confined to top scholars. It is as if the Reformation slowed down the mobility of average scholars. This is confirmed by the increase in the coefficient associated with cost distance.

3.4 Sample selection

Our database does not include the universe of professors. This implies two sources of sample selection issues: (i) many obscure scholars are not included in the sample, and (ii) there is a considerable heterogeneity in the coverage of institutions. We assess whether our results are robust to sample selection. Results are reported in col. (2) and col. (3) of Table 6. We report our benchmark results in col. (1).

As far as scholars are concerned, some are included in the sample but data on their place of birth are missing. This is usually the case for less well known professors. Indeed, among the scholars with a known birthplace, 52% have a positive q_i. This proportion drops to 14% for those with an unknown birthplace. Hence, our sample is likely to overweight top-quality professors (high q_i) and underweight the less well known (low q_i). This is a limitation because the co-existence of professors who are famous and those who are not is key to identifying selection and sorting patterns. To measure the importance of sample selection, we re-estimate the multinomial logit (5) by making the sample less selective. To do so, we use the identified scholars of unknown origin, and assume that they were born in the city of their university, implying $d_{ik} = \ln(\text{cost}^{\min})$ for them.

Col. (2) shows the results obtained when assuming that all identified scholars from unknown origin are locals. This increases the sample size by one third. Assuming scholars with unknown birthplace are locals reinforces substantially the positive selection effect, doubling the corresponding coefficient; the agglomeration terms and positive sorting are not much affected. These results suggest that if we had observed the whole universe of scholars, which contains many more unknown people born locally, positive selection would appear stronger while leaving sorting unaffected. Hence, our benchmark estimates likely give a lower bound on selection.

As far as institutions are concerned, we restrict our working sample to universities with at least 20 scholars (instead of 10 in the benchmark) in col. (3). This reduces a little the total number of observations, but reduces also the choice set of every scholar by removing some small universities. Our empirical results are highly robust to these changes. Finally, we also remove universities that are not well covered by our sources, i.e. those having only partial coverage (indicated by Cov= 1 in Tables 1). This amounts to remove 56 institutions from the set of 147 universities. Results are displayed in col. (4). We loose the significance of the communal freedom variable D_{kt}, probably because some important cities for identifying this effect were removed, but selection and sorting mechanisms are reinforced.

	Benchm	Sample Unknown	Coverage ≥ 20	Removing partial cov.
	(1)	(2)	(3)	(4)
d_{ik}	-1.752***	-1.924***	-1.756***	-1.828***
	(0.009)	(0.008)	(0.009)	(0.011)
$Q_{ik\tau(i)}$	0.184***	0.189***	0.184***	0.169***
	(0.007)	(0.006)	(0.007)	(0.007)
P_{kt}	0.168***	0.152***	0.168***	0.197***
	(0.023)	(0.022)	(0.023)	(0.025)
D_{kt}	0.142***	0.177***	0.145***	0.042
	(0.039)	(0.038)	(0.039)	(0.043)
$d_{ik}q_i$	0.030***	0.062***	0.030***	0.039***
	(0.003)	(0.003)	(0.003)	(0.003)
$Q_{ik\tau(i)}q_i$	0.015***	0.016***	0.016***	0.022***
	(0.001)	(0.002)	(0.001)	(0.002)
FE	yes	yes	yes	yes
N. Obs.	27,145	36,858	26,880	23,988
Log Likelihood	-58,999	-60,618	-57,803	-45,544

Table 6: Multinomial logit regressions: robustness to selection and coverage

3.5 Treatment of repeat movers

Remember that 10.9% of our scholars are linked to more than one university and we count each dyadic match as one observation. This raises two potential issues. Firstly, the weight of repeat movers exceeds that of one-time movers. As the number of career spells increases with human capital, this reinforces the over-representation of renowned scholars in our database. Secondly, by assuming that career-spell-specific choices are independent, we ignore the possibility that movers may have had correlated preferences.

The first problem can be easily addressed by removing repeat movers from the regression sample, which eliminates many famous scholars, or by linking them to a single university. We do both and, when following the second option, randomly select one of their affiliations. Solving the problem of correlated career spells is more complicated. To account for it, we generalize the standard logit model by relaxing the hypothesis of independence of individual choices. The independence property can be unrealistic in many settings, especially in situations with repeated choices over time. We can expect unobserved factors that affect a decision maker to persist over time. In a multinomial logit, we cannot include individual fixed effects since they would not affect the probability that a university k dominates another university k'. A more general deterministic component of utility can be written $V_{ik\tau(i)} = \beta_i \mathbf{x}_{ik\tau(i)}$, where β_i is a vector of coefficients that is unobserved for each i and varies randomly across professors, representing

	Benchm	Removing repeat movers	Repeat movers linked to 1 univ.	Mixed logit	Nested logit
	(1)	(2)	(3)	(4)	(5)
d_{ik}	-1.752***	-1.896***	-1.841***	-1.944***	-1.412***
	(0.009)	(0.011)	(0.011)	(0.013)	(0.015)
$Q_{ik\tau(i)}$	0.184***	0.190***	0.164***	0.189***	0.136***
	(0.007)	(0.008)	(0.008)	(0.008)	(0.006)
P_{kt}	0.168***	0.192***	0.170***	0.171***	0.152***
	(0.023)	(0.028)	(0.026)	(0.024)	(0.018)
D_{kt}	0.142***	0.163***	0.144***	0.197***	0.131***
	(0.039)	(0.046)	(0.044)	(0.041)	(0.029)
$d_{ik}q_i$	0.030***	-0.011**	0.014**	0.030***	0.029***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)
$Q_{ik\tau(i)}q_i$	0.015***	0.028***	0.023***	0.019***	0.013***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
ψ_1 (low)					0.795***
					(0.018)
ψ_2					0.760***
					(0.014)
ψ_3					0.774***
					(0.012)
ψ_4 (high)					0.719***
					(0.009)
FE	yes	yes	yes	yes	yes
N. Obs.	27,145	21,061	22,488	27,145	27,145
Log Likelihood	-58,999	-37,452	-42,568	-58,526	-58,693

Notes. In the mixed logit, the six variance parameters are estimated as well, four of them exhibit a variance that significantly differs from zero (variance of the coefficients of d_{ik}, D_{kt}, $d_{ik}q_i$, and $Q_{kt}q_i$).

Table 7: Robustness to Repeat Movers, Mixed and Nested logit

their tastes. This specification is the same as for the logit except that now the coefficients β_i vary in the population rather than being fixed. In particular, the coefficient β_i can be expressed as the sum of a population mean, $\overline{\beta}$, and an individual deviation, η_i, such that their utility of moving to destination k is written $U_{isk\tau(i)} = \overline{\beta}\mathbf{x}_{ik\tau(i)} + \eta_i\mathbf{x}_{ik\tau(i)} + \epsilon_{isk}$. The last two terms of such a *Random-Parameter Logit* capture the unobserved portion of utility. In other words, the marginal effect on the latent dependent variable is individual specific. The same tastes are used by the decision maker for each career spell and the variance in β_i induces correlation in utility across destinations and career spells.

How these parameters vary across individuals is unknown. The mixed logit model assumes that these parameters vary according to the population PDF $g(\beta_i|\theta)$, where θ represents the moments of the distribution such as the mean and the variance, which must be estimated. A fully parametric mixed logit model arises once $g(\beta_i|\theta)$ is specified. We assume that the coefficient vector is independent and normally distributed, $\beta_i \rightsquigarrow N\left(\overline{\beta}, \sigma^2\right)$. The unobserved portion of utility is correlated across destinations and career stages due to the common influence of η_i, which violates the IIA property of the standard logit (Revelt and Train 1998). The full parametric model can be estimated using the simulated maximum-likelihood procedure (Sarrias et al. 2016).

In col. (2) of Table 6, we show that most of our results are highly robust to the exclusion of repeat movers. Compared to the benchmark specification of col. (1), removing repeat movers slightly increases the magnitude of the agglomeration and gravity terms. As for sorting and selection, their identification relies on the difference between famous and obscure scholars in the sensitivity of location choices to institutional quality and distance. Remember repeat movers exhibit an average ability index that is 2.3 times greater than the mean; they account for 10.9% of our scholars and are linked to 2.4 universities, on average. Removing them from the sample decreases the number of observed dyads by 22.5% (from 27,145 to 21,061) and eliminates many famous scholars at the upper end of the ability distribution. The sorting term resists this change and its magnitude is drastically strengthened compared to the benchmark. This is important because we will see it is the effect that is driving our simulation results in Section 4. By contrast, the selection term turns negative and becomes less significant. Instead of removing entirely the repeat movers, we keep them but associate them with only one of their affiliations (randomly chosen) in col. (3). The selection term becomes positive again and is equal to half of the level obtained in the benchmark regression. This demonstrates once again that including famous and obscure scholars is key to identifying sorting and selection patterns. In particular, the significance and the magnitude of the selection term are strongly governed by the fact that the location choices of (high-ability) repeat movers are less sensitive to distance than those of lower-ability scholars.

In col. (4) of Table 7, we relax the assumption of independent career choices for multi-destination scholars, and estimate a mixed logit model with individual-specific vectors of coefficients drawn from a normal distribution. The agglomeration, selection, and sorting mechanisms are preserved. Although the mixed logit entails six additional parameters (the s.e. of the six coefficients - not reported), a likelihood ratio test would reject the benchmark formulation in favor of the mixed logit formulation. The mixed logit has a disadvantage though: the estimates are obtained by simulation, while in the multinomial logit, a likelihood function is maximized. In addition, the results depend on the assumption regarding the distribution of the random parameters.

3.6 Endogeneity of $\mathbf{q}_i$

The most problematic endogeneity issue arises because the ability of each professor i is measured by an index of human capital observed a long time after the end of her career (q_i), which is likely to be influenced by the quality of the university that was chosen. This means that we should distinguish between $\overline{q}_i$, the innate/exogenous level of ability, and q_i, the ex post level of notability. Let us denote by k^* the university chosen by a scholar. Ideally, we should use $\overline{q}_i$ to estimate the multinomial logit (5). However, we only observe q_i, and this ex post level might be affected by $Q_{k^*\tau(i)}$, the notability of the chosen university. This implies that we do not observe the potential level of human capital if the individual had been working at a different university k. Assume for example that $q_i = \overline{q}_i + \theta Q_{k^*\tau(i)}$ and denote by $\overline{V}_{ik\tau(i)}$ the indirect utility level obtained after replacing q_i by $\overline{q}_i$ in (9).

In theory, the multinomial logit implies that university k dominates university k' if $\overline{V}_{ik\tau(i)} + \epsilon_{isk} > \overline{V}_{ik'\tau(i)} + \epsilon_{isk'}$, which only depends on the characteristics of individual i and universities k and k'. In practice, we are unable to model $\overline{V}_{ik\tau(i)}$ and $\overline{V}_{ik'\tau(i)}$ properly because our measure of individual human capital is k^*-specific (i.e. influenced by $Q_{k^*\tau(i)}$). The endogeneity of q_i implies that the difference in utility is measured with additional noise: university k dominates university k' if

$$\overline{V}_{ik\tau(i)} + \epsilon_{isk} > \overline{V}_{ik'\tau(i)} + \epsilon_{isk'} + \theta Q_{k^*\tau(i)} \Delta_{ikk'\tau(i)}, \tag{10}$$

where $\Delta_{ikk'\tau(i)} \equiv \beta_4 \left(Q_{ik\tau(i)} - Q_{ik'\tau(i)} \right) + \beta_6 \left(d_{ik} - d_{ik'} \right)$ results from the two interaction terms that are affected by our noisy measure of individual human capital in Eq. (9). The term $+\theta Q_{k^*\tau(i)} \Delta_{ikk'\tau(i)}$ in (10) is correlated across destinations, due to the presence of $Q_{k^*\tau(i)}$. Hence, the inability to observe $\overline{q}_i$ leads to the violation of the IIA property.

To mitigate this problem, we estimate a nested logit model (McFadden 1978) where nests are defined as groups of universities sharing similar levels of notability ($Q_{ik\tau(i)} \approx Q_{ik'\tau(i)}$) during the period of activity of individual i. We partition the choice set $K_{\tau(i)}$ into four groups of alternatives, $K_{m\tau(i)}$ with $m = (1, 2, 3, 4)$ for the top, mid-high, mid-low, and bottom universities. Our partition is based on the notability index observed in the 4th and 5th periods. Each university belongs to exactly one nest. Building on Ortega and Peri (2013) and Bertoli and Moraga (2015), we assume that the individual random taste shock is a mixture of a location-specific and a nest-specific term:

$$\epsilon_{isk} = \psi_m \upsilon_{isk} + (1 - \psi_m)\upsilon_{im},$$

where $\psi_m \in [0, 1]$ is the weight associated with the location-specific term, υ_{isk}, which is assumed to be independently and identically distributed as EVT-I; and υ_{im} is an error term that is specific to the m^{th} nest ($k \in K_{m\tau(i)}$), whose distribution depends on ψ such that the marginal distribution of ϵ_{isk} is also EVT-I (Cardell 1997). Parameter ψ_m also determines the mutual correlation in the realizations of the nest-specific error term. We have $\psi_m = \sqrt{1 - \rho_m}$, where ρ_m represents the correlation coefficient within nest m. Hence, ψ_m is a dissimilarity parameter. The higher ψ_m, the smaller the weight of the nest-specific component and the smaller the within-nest correlation of error term. When $\psi_m = 1$ for all m, the nested logit boils down to the standard multinomial logit ($\epsilon_{isk} = \upsilon_{isk}$).

The nested logit model assumes a generalized version of the EVT-I distribution, such that (i) the mean error varies across nests, and (ii) alternatives within a nest exhibit mutually correlated error terms (but the same mean). On the contrary, the error terms of two alternatives belonging to different nests are uncorrelated but have different means. In our context, this difference in the means captures the component of the error term $\theta Q_{ik^*\tau(i)} \Delta_{ikk'\tau(i)}$ and hence corrects for the endogeneity bias. It reflects the influence of the chosen university on individual quality. Within a nest, this component is close to zero because $\Delta_{ikk'\tau(i)} \approx 0$. Notice that this technique to correct for the endogeneity bias is possible only because the q_i always appears interacted with a variable for which we can build nests, and never appears alone (it cannot explain location choice alone as it is not destination specific).

The probability of individual i choosing university $k \in K_{m\tau(i)}$ is equal to the product of the probability of choosing alternatives in nest $K_{m\tau(i)}$ and the probability of choosing exactly k in $K_{m\tau(i)}$ (Heiss et al. 2002). It is given by

$$p_{isk\tau(i)} = \frac{\exp(\beta \mathbf{x}_{ik\tau(i)}/\psi_m)}{\exp(IV_{m\tau(i)})} \times \frac{\exp(IV_{m\tau(i)}\psi_m)}{\sum_{m'} \exp(IV_{m'\tau(i)}\psi_m)} \quad \forall t, \tag{11}$$

where $IV_{m\tau(i)} = \ln \sum_{k' \in K_{m\tau(i)}} \exp(\beta \mathbf{x}_{ik\tau(i)}/\psi_m)$ is the inclusive value of each nest $K_{m\tau(i)}$, representing the rescaled measure of attractiveness of the nest for individual i (i.e. the expected value of the utility individual i obtains from the alternatives in nest $K_{m\tau(i)}$).

Finally, in col. (5) of Table 7, we estimate the nested logit model to mitigate endogeneity concerns about q_i. Our nests are defined as groups of universities sharing similar levels of notability. Compared to the benchmark, the effects of agglomeration are weakened but still positive and highly significant; part of the agglomeration force is likely to be captured by the nest-specific error term. The selection coefficient is slightly greater. Sorting is weakened but remains highly significant and important compared to the agglomeration effect: when q_i is around 10, positive sorting increases the gains from settling in a more prestigious university by a factor of 2. Note that we reject the assumption of no nests, either through a likelihood ratio test (lr=611.413), or by testing whether the correlations within nests are zero, or equivalently $\psi_m = 1\ \forall n$ (Wald=1062.255). We also reject that the degree of correlation inside each nest is the same, $\psi_m = \psi_{\tau(i)}\ \forall m$ (Wald=59.158, p-val=0.000).

3.7 Endogeneity of $Q_{k\tau(i)}$

We close the discussion on econometric issues with a few words on the potential endogeneity of $Q_{k\tau(i)}$. In our benchmark model, we aim to identify the effect of $Q_{k\tau(i)}$ on the probability that scholar i chooses university k in period τ. One might fear a second endogeneity problem arising from the fact that the performance of each university k is measured by an ex post index of notability, which is determined by the quality of some professors having chosen to locate there. We do not think this is a major issue in our context for several reasons. Firstly, remember that we replace $Q_{k\tau(i)}$ by $Q_{ik\tau(i)}$ in our estimations; the latter is computed after excluding individual i from the notability index. Secondly, top scholars might still have an effect on their peers. We define the notability of an institution in a given period as a function of the notability of its top five professors. Over the whole period under consideration, our estimation will rely on 147 universities and 23,624 scholars with known birthplaces making 27,145 location choices. The notability over the eight periods of our 147 institutions depends on the quality of 2,932 professors (i.e. 10.8% of our sample only). The risk of a reverse causal impact of q_i (and thus $p_{isk\tau(i)}$) on $Q_{ik\tau(i)}$ is less of an issue for 89% of our scholars. Thirdly, in Eq. (3), we allow for some persistency of past notability, which allows to assess the risk of endogeneity by letting the persistency parameter vary. Finally, notability is computed over periods which are longer than a human life, implying that the top scholars at a university for a given period did not necessarily meet in person, which also limits the endogeneity coming from face-to-face interactions. Hence, we consider $Q_{ik\tau(i)}$ as exogenous in our empirical analysis.

4 Implications for the Scientific Revolution

Firstly, we assess the estimated effect of the determinants of location decisions (x_{ikt}) on individual choice probabilities by comparing simulated outcomes with counterfactual experiments. We focus on the role of agglomeration, positive selection, and positive sorting for two scholars who were born in the same region in the first period (the period with the smallest choice set), but who exhibit different levels of human capital. We first consider Thomas Aquinas (1225-1274), who was born in Roccasecca and taught theology at Paris for twenty years (1252-1272) and at Naples. Thomas Aquinas belongs to the very top of the ability distribution (ranked 9th, $q = 11.66$). Aquinas's choice set consists of 24 universities. We then consider Roffredus Beneventanus (1170-1243), who was born in Benevento, taught law at Naples (1170-1243), and is in the middle of the distribution ($q = 2.83$). As he was born 50 years before Thomas Aquinas, Roffredus Beneventanus's choice set consists of just 20 universities. In Table 8, we compare the predicted location probabilities of the full specification of Table 4, with those of a restricted model in which the coefficients of the selection and sorting terms are set to zero. For both scholars, the set of universities is ranked by decreasing order of the predicted probability generated by the full specification.

Unsurprisingly, positive selection implies a broader menu of effective options, whereas agglomeration forces and positive sorting increase the attractiveness of famous universities. The effects are balanced by the extent of each scholar's notability. Our estimated model shows that the most likely locations for Thomas Aquinas are Bologna (32.1%), Naples (25.8%), Padova (10.3%) and Paris (8.7%). Neutralizing the positive selection term increases the probability of choosing a good location closest to his birthplace (Naples) to the detriment of Bologna and Paris. Neutralizing the positive sorting and agglomeration term drastically decreases the attractiveness of Paris. Overall, the basic gravity model predicts a low probability of choosing Paris (0.6%). The combination of agglomeration, selection, and positive sorting increases this probability by a factor of 14, and increases the probability of choosing Bologna by a factor of 5.

For Roffredus Beneventanus, who has less human capital (q_i), the selection and sorting effects are weaker. Our estimated model shows that the most likely location is Naples (34.9%). Compared to Thomas Aquinas, Roffredus Beneventanus is more sensitive to distance, and less sensitive to the notability of the university or to agglomeration effects. Removing the sorting effect or the agglomeration effect increases the probability of choosing Padua, Salerno, and Pisa, at the expense of Bologna, Montpellier, and Paris. When removing the selection effect, similar changes are obtained. The basic gravity model also predicts that Naples and Padua are the two most attractive universities, for both Roffredus Beneventanus and Thomas Aquinas, and that the probability of choosing Paris or Bologna, the best universities in this period, are similar for both (a bit higher in fact for Beneventanus, because his choice set is more limited. However, unlike for Thomas Aquinas, the combination of agglomeration, positive selection, and positive sorting increases the probability of choosing Paris and Bologna by a factor of 5 to 3 only.

Secondly, we go beyond individual cases by using our estimated model to simulate the contribution of agglomeration, positive selection, and positive sorting to total university output by period. We construct a proxy for the total output of university k for period τ, denoted by $Y_{k\tau}$, which is an aggregation of the human capital of all scholars predicted to work there. $Y_{k\tau}$ is defined as a CES (constant elasticity of substitution) combination of the ability levels of its predicted members:

$$Y_{k\tau} = \left(\sum_i \hat{p}_{ik\tau(i)} q_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{12}$$

where $\hat{p}_{ik\tau(i)}$ is the weight given to professor i at university k in period $\tau(i)$. We set it equal to the simulated probability that i goes to k from the multinomial logit model – like the probabilities shown in Table 8 for two cases. Parameter σ represents the elasticity of substitution between academic scholars' human capital in production.

	Estimated	No agglom	No selec.	No sorting	Basic Gravity
			Thomas Aquinas (K_i= 24)		
Ubologna-1088	32.1%	17.1%	27.4%	19.1%	6.8%
Unapoli-1224	25.8%	29.0%	36.9%	30.5%	38.4%
Upadua-1222	10.3%	14.6%	9.1%	14.2%	13.9%
Uparis-1200	8.7%	2.9%	5.1%	4.0%	0.6%
Usalerno-1231	6.9%	8.2%	8.8%	7.5%	8.8%
Umontpellier-1289	3.6%	2.9%	2.6%	2.7%	1.2%
Upisa-1343	3.5%	8.5%	3.3%	7.8%	13.6%
Others (17)	9.1%	16.8%	6.8%	14.2%	16.6%
			Beneventanus Roffredus (K_i= 20)		
Unapoli-1224	35.0%	34.9%	36.4%	35.9%	36.1%
Usalerno-1231	28.3%	29.6%	30.6%	28.4%	31.2%
Ubologna-1088	15.4%	7.2%	14.0%	13.4%	5.6%
Upadua-1222	9.5%	12.0%	8.8%	10.2%	11.4%
Upisa-1343	3.5%	7.4%	3.2%	4.2%	7.9%
Uparis-1200	2.5%	0.7%	2.1%	2.0%	0.5%
Umontpellier-1289	1.6%	1.1%	1.4%	1.5%	0.9%
Others (13)	4.2%	7.0%	3.7%	4.5%	6.5%

Table 8: Role of selection and sorting: two examples

The simulated output $Y_{k\tau}$ should be interpreted as including advancement to knowledge, quality of teaching, and service to society (such as supplying cautious physicians, rigorous lawyers to the local courts, or well-educated priests and pastors to parishes). Then, we compare the total simulated output

$$Y_\tau = \sum_k Y_{k\tau}$$

in the benchmark (normalized to 100) with the predicted output obtained after neutralizing the agglomeration, positive selection, and sorting terms. Hence, the important point here is not the level of Y_τ in itself, but the gap between Y_τ with and without market forces.

When the elasticity of substitution tends to infinity ($\sigma = \infty$), we have perfect substitutability between scholars. The total output is the sum of individual human capital, independent of location (represented by the $\hat{p}_{ik\tau(i)}$):

$$\lim_{\sigma \to \infty} Y_\tau = \sum_k \sum_i \hat{p}_{ik\tau(i)}\, q_i = \sum_i q_i.$$

By contrast, when σ is finite, there is a complementarity relationship between academic scholars. The smaller σ, the greater the knowledge gain from agglomerating high-ability scholars at the same university, and the agglomeration of the highest ability scholars leads to output gains. In our benchmark regressions and simulations, we use a CES production function with $\sigma = 2$, in line with the definition of the notability of the university in Eq. (3).

Table 9 shows the results obtained with the benchmark model (i.e. standard multinomial logit model with $\sigma = 2$) in the top panel, with the nested logit variant in the middle panel, and with the high-complementarity variant (i.e. standard multinomial logit model with $\sigma = 1.2$) in the bottom panel. In the nested logit variant, we use the estimates provided in col. (4) of Table 7. The high-complementarity variant relies on col. (3) of Table 2. In each panel, col. (2) to (4) give the total output obtained after neutralizing the effect of agglomeration, positive selection, and positive sorting. In col. (5), we neutralize these three mechanisms jointly, while keeping the distance term and the university fixed effects (i.e. basic gravity).

In line with our empirical findings, col. (3) of Table 9 shows that positive selection hardly influences the total simulated output. We have seen in Table 8 that positive selection tends to scatter talents across universities by increasing the menu of options for the highest ability scholars, but the total effect of this increased dispersion is small. On the other hand, cols. (2) and (4) show that agglomeration and positive sorting play an important role, especially in the earlier periods, when there are few universities. Under the benchmark and nested-logit variants, agglomeration and sorting increase the total output of Europe by about 55% before the Black Death (32 universities), by about 35% before the invention of the printing press (49 universities), and by 40% before the rise of Protestantism (74 universities).

It is worth noticing that sorting and agglomeration do not necessarily increase total simulated output. Their joint effect on output depends on the correlation between the notability of universities and the level of city/university amenities (captured, in our regressions, by the university fixed effect and the attractiveness of the city). When the correlation is high, the effects of notability and amenities point in the same direction; the best scholars agglomerate in the best universities. When the correlation is lower, agglomeration and sorting can result in the concentration of talent in second-best universities, which reduces total academic output. This is at least the case if the intensity of agglomeration and sorting forces is limited.

The effects of agglomeration and sorting become weaker after 1618. They increase total output by about 20% over this period. Overall, we find that agglomeration and sorting are more likely to reduce academic output when the individual choice set is large. For this reason, their effect on academic output diminished in the late 16th and in the 17th centuries, when the number of universities almost doubled. What is specific to periods 5-7 is the presence of many universities with a large number of scholars having published something ($q_i > 0$) which was not highly influential. Shutting down agglomeration redistributes superstars to the advantage of these less prestigious universities, thus increasing the level of the many average people there.

To further understand the role of market forces, Figure 2 maps the winners (in green) and losers (in red) due to market forces in the period 1618-1685. The surface of each circle represents the difference in simulated output between the benchmark case and the basic gravity case.

	No agglom.	No selection	No Sorting	Basic Gravity
		Benchmark ($\sigma = 2$)		
1000-1199	76.5	97.6	90.0	73.7
1200-1347	50.7	97.3	77.9	45.1
1348-1449	70.1	97.4	90.1	64.9
1450-1526	72.6	98.5	92.3	66.4
1527-1617	67.6	96.6	89.5	60.4
1618-1685	78.5	98.1	92.4	74.6
1686-1733	83.4	98.9	94.5	80.1
1734-1800	86.6	98.8	94.6	83.1
		Nested logit ($\sigma = 2$)		
1000-1199	76.2	96.8	88.5	72.3
1200-1347	51.2	96.9	75.7	44.3
1348-1449	71.0	97.0	89.2	65.1
1450-1526	74.6	98.1	91.9	68.0
1527-1617	69.8	96.2	88.7	62.2
1618-1685	81.7	98.2	92.3	77.2
1686-1733	85.7	99.0	94.3	81.7
1734-1800	87.6	98.6	93.8	83.2
		High complementarity ($\sigma = 1.2$)		
1000-1199	42.3	96.0	96.5	39.4
1200-1347	34.9	95.8	93.9	32.1
1348-1449	36.7	98.3	94.5	34.3
1450-1526	41.3	98.8	94.8	39.1
1527-1617	34.9	96.8	92.8	31.9
1618-1685	37.6	97.0	93.1	34.6
1686-1733	36.4	97.1	93.5	33.5
1734-1800	35.7	97.2	92.6	32.5

Table 9: Effect of agglomeration, selection, and sorting on academic output (100=benchmark)

An easily understandable case is Lund vs Copenhagen. With market forces, scholars born in Sweden are more likely to locate in Copenhagen which has a high notability ($Q_6 = 7.35$) rather than in Lund, which is just average ($Q_6 = 4.25$), while without agglomeration and sorting forces, they are content with Lund. We also note that Rinteln is a big loser in Germany, being surrounded by many good universities such as Leipzig and Jena. It is also noteworthy that the South of Europe is not doing so poorly, although the bigger gains are in the North. Renowned Southern universities still attract talents (Salamanca, Padua, Bologna, and Rome). In a sense, without market forces, the fate of the South would have been worse.

When looking at universities which were permanently closed down over the period 1700-1900, many of them were losers from the market in the last 3 periods. Altdorf (closed in 1809), Bamberg (closed in 1803), Cahors (closed in 1751), Cervera (closed in 1821), Dorpat (closed in 1710), Harderwijk (closed in 1811), Pont-à-Mousson (closed in 1768), Rinteln (closed in 1809), Siguenza (closed in 1807), and Valence (closed in 1793) are in this category. A few winners were closed too: Erfurt (closed in 1816) and Frankfurt (Oder) (closed in 1811).

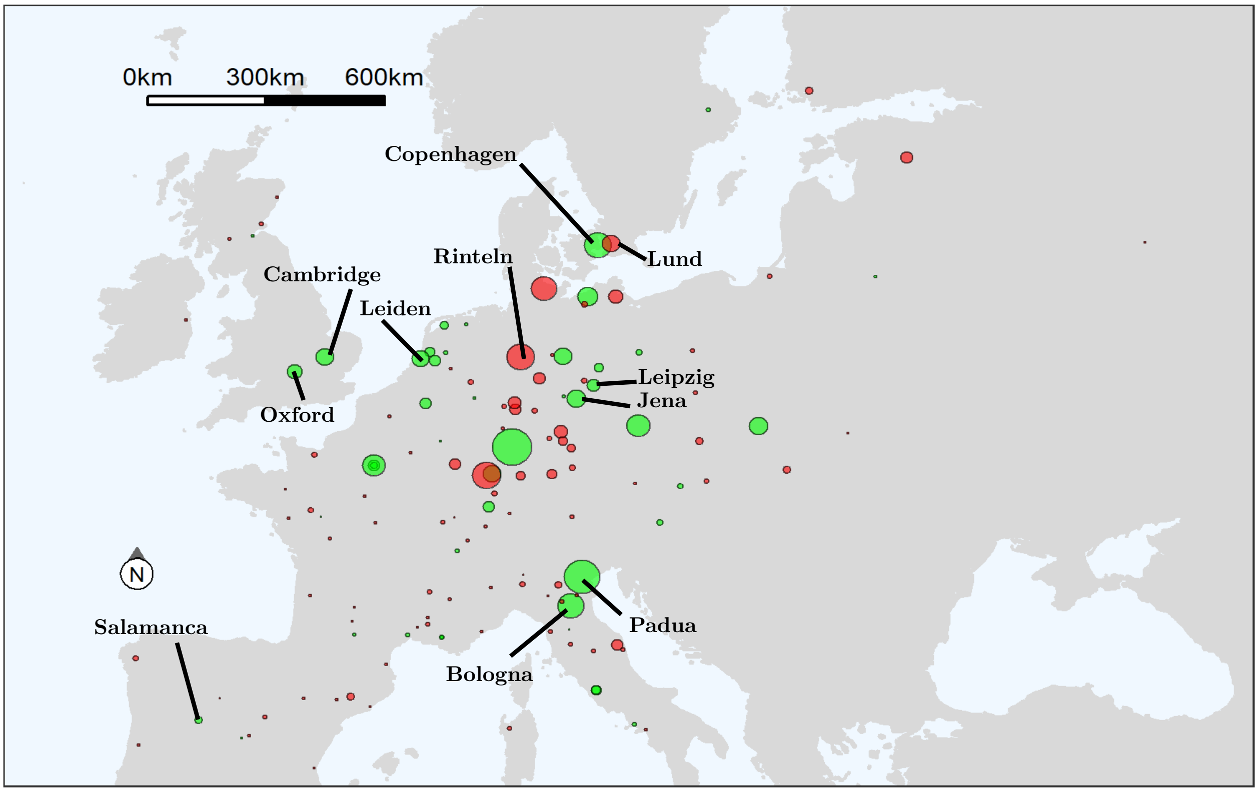


Figure 2: Winners (green) and Losers (red) from Market Forces in 1618-1685

Overall, our results show that agglomeration and sorting effects in the academic market contribute to fostering university output. The size of the agglomeration and sorting effects before the middle of the 16th century are quantitatively significant. Without these effects, university output would be reduced by 50%. As we do not model any cumulative effect of knowledge creation, this 50% should be understood as a lower bound.

Several economic historians claim that labor markets were relatively complete and competitive in Medieval Europe: “Given the low reproductive success of the urban population there had to be a constant flow of labor from the country to the city (Clark 2008). The records of a 1292 tax levied by Philip the Fair on the commoner households of Paris show that 6 percent were foreigners: 2.1 percent English, 1.4 percent Italian, 0.8 percent German, 0.7 percent Flemish, 0.6 percent Jewish, and 0.4 percent Scottish" (Clark 2008, Sussman 2006). We can compare these numbers with the origin of the scholars of Paris University in the first two periods of our sample (1000-1347). Based on the 394 persons with known origin, we obtain that 54% of the scholars were born in France (in its 2020 limits), 21% are British, 4% are from Germany, 10% are from Italy, and 5% are from the Low Countries – the data for this period are mostly based on Courtenay (1999), Gorochov (2012), and Genet (2019). Although the mobility of ordinary people seems quite high already, the mobility of university scholars is higher by an order of magnitude.

The importance of market forces seem particularly relevant in the periods preceding and coinciding with the dawn of the Scientific Revolution, a period commonly defined as spanning Copernicus's and Newton's times, i.e. 1543-1687 (Applebaum 2003). In the last two centuries before the Industrial Revolution, these effects decrease significantly or even become non-existent. Hence, although we provide no causal evidence of such a link, our simulations lend credence to the hypothesis that universities might have been key to triggering the rise of this new science. This view is corroborated by the analysis of the gains from the market at the local level. In our simulations, the universities gaining the most from agglomeration and sorting forces in the period 1450-1526 are Rome, Bologna, Padua, Paris, and Louvain. In the period 1527-1617, one can add Cambridge and Leiden to the list. Those were indeed leading universities for the Scientific Revolution.

Our results are conditional on the value of σ. As stated above, all these effects would disappear if we had taken $\sigma = \infty$, as the allocation of scholars (represented by the $\hat{p}_{ik\tau(i)}$) across places would not matter. Although the benefits of sorting however only slightly exceed those obtained in the benchmark, the positive effects of agglomeration are magnified under the high-complementarity variant with $\sigma = 1.2$. In addition, contrary to the benchmark, the effect of agglomeration is relatively stable across periods. This means that the huge gains from agglomeration in the top universities are not compensated for by losses in average- and low-quality universities in the more recent periods. Overall, when combining all mechanisms, the simulated output increases threefold in most periods.

5 Conclusions

In European universities, students were educated by a plurality of masters, and schools were open to students and scholars from all parts of Europe. In this paper, we map the European academic market in the medieval and early modern times. We build an original database of thousands of scholars from university sources to study the location pattern of scholars over the period 1000-1800. The quality of scholars is measured using information provided by Worldcat and Wikipedia. Using a multinomial logit, we show that scholars tend to agglomerate in the best universities, and that this phenomenon is more pronounced within the upper tail of the talent distribution: better scholars are more sensitive to the quality of the university (positive sorting), and migrate over greater distances (positive selection). Agglomeration and sorting patterns influenced the distribution of upper-tail human capital across Europe, and contributed to fostering university output at the dawn of the Scientific Revolution.

Agglomeration, sorting, and selection testify to market forces at work. They appear when there is a competition between universities to attract scholars, or among scholars to land the best positions available. This contrasts with a common but mistaken view that markets are a modern phenomenon, but our findings are in line with the qualitative evidence put forward by historians such as Denley (2013) who describes the emergence in Italy of "an efficient and sometimes cut-throat academic market, with its own 'transfer season,' clearly defined hierarchies, rocketing salaries for the top players, and a mentality of academic celebrity that fed off it." At the European level, two features may have helped the academic market to develop. First, political fragmentation, together with competition between church and state, prevented a centralized control by the political sphere of universities. Second, the use of Latin as a *lingua franca*, which persisted late into the early modern period, allowed scholars to teach anywhere at low cost.

Our simulations suggest that the presence of a functioning academic market in Europe helped universities to produce more at the dawn of European primacy. This might have paved the way for the Enlightenment, humanistic, and scientific revolutions. We thus provide some quantitative support to the views developed by historians, such as Huff (2017)'s approach to the Scientific Revolution, comparing the West to China and the Islamic World. Huff suggests that the origins of the stronger support given to scientific inquiry in the West during the early modern period can be traced back to the medieval period when European institutions were reconstructed. In this context, he sees the rise of European universities in the Middle Ages and their long-run contribution to the Scientific Revolution as highly significant.

References

Abowd, John, Francis Kramarz, and David Margolis. 1999. "High wage workers and high wage firms." *Econometrica* 67 (2): 251–333.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration." *American Economic Review* 102 (5): 1832–56.

———. 2014. "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration." *Journal of Political Economy* 122 (3): 467–506.

Abramitzky, Ran, and Fabio Braggion. 2006. "Migration and human capital: self-selection of indentured servants to the Americas." *Journal of Economic History* 66 (4): 882–905.

Agrawal, Ajay, Devesh Kapur, John McHale, and Alexander Oettl. 2011. "Brain drain or brain bank? The impact of skilled emigration on poor-country innovation." *Journal of Urban Economics* 69 (1): 43–55.

Akcigit, Ufuk, Salomé Baslandze, and Stefanie Stantcheva. 2016. "Taxation and the international mobility of inventors." *American Economic Review* 106 (10): 2930–2981.

Applebaum, Wilbur. 2003. *Encyclopedia of the scientific revolution: from Copernicus to Newton.* New York: Routledge.

Arabeyre, Patrick, Jean-Louis Halpérin, and Jacques Krynen. 2007. *Dictionnaire historique des juristes français (XIIe-XXe siècle).* Paris: Presses Universitaires de France.

Bairoch, Paul, Jean Batou, and Pierre Chevre. 1988. *Population des villes européennes de 800 à 1850: banque de données et analyse sommaire des résultats.* Librairie Droz.

Beine, Michel, Frederic Docquier, and Caglar Ozden. 2011. "Diasporas." *Journal of Development Economics* 95 (1): 30–41.

Beltrán Tapia, Francisco J, and Santiago de Miguel Salanova. 2017. "Migrants self-selection in the early stages of modern economic growth, Spain (1880–1930)." *Economic History Review* 70 (1): 101–121.

Bertoli, Simone, and Jesús Fernández-Huertas Moraga. 2015. "The size of the cliff at the border." *Regional Science and Urban Economics* 51:1–6.

Bertoli, Simone, and Ilse Ruyssen. 2018. "Networks and migrants' intended destination." *Journal of Economic Geography* 18 (4): 705–728 (03).

Bogart, Dan. 2011. "Did the Glorious Revolution contribute to the transport revolution? Evidence from investment in roads and rivers 1." *Economic History Review* 64 (4): 1073–1112.

Borjas, George J. 1987. "Self-selection and the earnings of immigrants." *American Economic Review* 77 (4): 531–553.

Bosker, Maarten, Eltjo Buringh, and Jan Luiten Van Zanden. 2013. "From Baghdad to London: Unraveling urban development in Europe, the Middle East, and North Africa, 800–1800." *Review of Economics and Statistics* 95 (4): 1418–1437.

Bourchenin, Pierre Daniel. 1882. *Étude sur les académies protestantes en France au XVIe et au XVIIe siècle*. Grassart.

Breschi, Stefano, and Francesco Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of Economic Geography* 9 (4): 439–468.

Buonanno, Paolo, Matteo Cervellati, Sara Lazzaroni, and Giovanni Prarolo. 2019. "Political Institutions, Culture and Cooperation: Medieval Social Contracts and their Legacy." Mimeo, University of Bologna.

Cantoni, Davide, Jeremiah Dittmar, and Noam Yuchtman. 2018. "Religious competition and reallocation: the political economy of secularization in the Protestant Reformation." *The Quarterly Journal of Economics* 133 (4): 2037–2096.

Cantoni, Davide, and Noam Yuchtman. 2014. "Medieval universities, legal institutions, and the commercial revolution." *Quarterly Journal of Economics* 129 (2): 823–887.

Cardell, N Scott. 1997. "Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity." *Econometric Theory* 13 (2): 185–213.

Chiquiar, Daniel, and Gordon H Hanson. 2005. "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113 (2): 239–281.

Cinnirella, Francesco, and Jochen Streb. 2017. "The role of human capital and innovation in economic development: evidence from post-Malthusian Prussia." *Journal of Economic Growth* 22 (2): 193–227.

Clark, Gregory. 2008. *A farewell to alms: a brief economic history of the world.* Princeton University Press.

Cottini, Elena, Paolo Ghinetti, and Simone Moriconi. 2019. "Higher Education Supply, Neighbourhood Effects and Economic Welfare." CESifo Working Paper.

Courtenay, William J. 1999. *Parisian scholars in the early fourteenth century: a social portrait.* Volume 41. Cambridge University Press.

Covarrubias, Matías, Jeanne Lafortune, and José Tessada. 2015. "Who Comes and Why? Determinants of Immigrants Skill Level in the Early XXth Century US." *Journal of Demographic Economics* 81 (1): 115–155.

Croissant, Yves. 2012. "Estimation of multinomial logit models in R: The mlogit Packages." available from R-Forge.

Dahl, Gordon B. 2002. "Mobility and the return to education: Testing a Roy model with multiple markets." *Econometrica* 70 (6): 2367–2420.

De la Croix, David, Matthias Doepke, and Joel Mokyr. 2018. "Clans, Guilds, and Markets: Apprenticeship Institutions and Growth in the Preindustrial Economy." *Quarterly Journal of Economics* 133 (1): 1–70.

De la Croix, David, and Omar Licandro. 2015. "The longevity of famous people from Hammurabi to Einstein." *Journal of Economic Growth* 20 (3): 263–303.

De la Croix, David, and Fabio Mariani. 2015. "From Polygyny to Serial Monogamy: A Unified Theory of Marriage Institutions." *Review of Economic Studies* 82 (2): 565–607.

Denley, Peter. 2013. "Medieval, Renaissance, modern. Issues of periodization in Italian university history." *Renaissance Studies* 27 (4): 487–503.

Deuster, Christoph, and Frédéric Docquier. 2019. "International Migration and Human Capital Inequality: A Dyadic Approach." In *Global Education Monitoring Report*, 1–41. UNESCO.

Dittmar, Jeremiah. 2019. "Economic Origins of Modern Science: Technology, Institutions, and Markets." London School of Economics.

Dittmar, Jeremiah, and Ralf Meisenzahl. 2019. "Public goods institutions, human capital, and growth: evidence from German history." *Review of Economic Studies*, 02.

Dowey, James. 2017. "Mind over matter: access to knowledge and the British Industrial Revolution." Ph.D. diss., The London School of Economics and Political Science (LSE).

Dulieu, Louis. 1979. *La médecine à Montpellier, vol II: La Renaissance.* Avignon: Les presses universelles.

Eloy, Nicolas FJ. 1755. *Dictionnaire Historique De La Médecine: Contenant Son Origine, Ses Progrès, ses Révolutíons, ses Sectes & son Etat chez différens Peuples.* Bassompierre.

Fink, Carsten, Ernest Miguelez, and Julio Raffo. 2013. "The global race for inventors." mimeo.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2012. "Foreign Born Scientists: Mobility Patterns for Sixteen Countries." *Nature Biotechnology* 30 (12): 1250–1253.

Frijhoff, Willem. 1996. "Patterns." Chapter 2 of *A History of the University in Europe. Vol. II: Universities in Early Modern Europe (15001800)*, edited by Hilde de Ridder-Symoens. Cambridge University Press.

Gallin, Joshua Hojvat. 2004. "Net migration and state labor market dynamics." *Journal of Labor Economics* 22 (1): 1–21.

Galor, Oded. 2011. *Unified growth theory.* Princeton University Press.

Galor, Oded, and Omer Moav. 2002. "Natural Selection and the Origin of Economic Growth." *Quarterly Journal of Economics* 117 (4): 1133–1191.

Genet, Jean-Philippe. 2019. "Projet Studium Parisiense." Université Paris 1.

Gergaud, Olivier, Morgane Laouenan, and Etienne Wasmer. 2017. "A Brief History of Human Time. Exploring a database of 'notable people'." mimeo.

Gibson, John, and David McKenzie. 2014. "Scientific mobility and knowledge networks in high emigration countries: Evidence from the Pacific." *Research Policy* 43 (9): 1486–1495.

Gorochov, Nathalie. 2012. *Naissance de luniversité. Les écoles de Paris dInnocent III à Thomas dAquin (v. 1200-v. 1245).* Paris: Honoré Champion.

Greif, Avner. 2006. "Family Structure, Institutions, and Growth: The Origins and Implications of Western Corporations." *American Economic Review* 96 (2): 308–312.

Grogger, Jeffrey, and Gordon H Hanson. 2011. "Income maximization and the selection and sorting of international migrants." *Journal of Development Economics* 95 (1): 42–57.

———. 2015. "Attracting talent: location choice of foreign-born PhDs in the United States." *Journal of Labor Economics* 33 (S1): 5–38.

Harris, Jonathan. 1995. *Greek emigres in the West 1400-1520.* Porphyrogenitus.

Heiss, Florian, et al. 2002. "Structural choice analysis with nested logit models." *Stata Journal* 2 (3): 227–252.

Herbermann, Charles George. 1913. *The Catholic Encyclopedia.* New York: Encyclopedia Press, Incorporated.

Hermans, Jos MM, and Marc Nelissen. 2005. *Charters of foundation and early documents of the universities of the Coimbra Group.* Leuven University Press.

Hoisl, Karin. 2007. "Tracing mobile inventorsthe causality between inventor mobility and inventor productivity." In *A Study of Inventors*, 65–119. Springer.

Huff, Toby E. 2017. *The rise of early modern science: Islam, China, and the West.* Cambridge University Press.

Jaeger, C Stephen. 2013. *The envy of angels: Cathedral schools and social ideals in Medieval Europe, 950-1200.* University of Pennsylvania Press.

Junius Institute. 2013. "Post Reformation Digital Library – Scholastica." `www.prdl.org`.

Keane, Michael P, and Kenneth I Wolpin. 1997. "The career decisions of young men." *Journal of Political Economy* 105 (3): 473–522.

Kedar, Benjamin. 1995. "Benvenutus Grapheus of Jerusalem, an Oculist in the Era of the Crusades." *Korot* 11:14–41.

Kerr, Sari Pekkala, William Kerr, Çağlar Özden, and Christopher Parsons. 2016. "Global talent flows." *Journal of Economic Perspectives* 30 (4): 83–106.

———. 2017. "High-skilled migration and agglomeration." *Annual Review of Economics* 9:201–234.

Kerr, Sari Pekkala, William R Kerr, and William F Lincoln. 2015a. "Firms and the economics of skilled immigration." *Innovation Policy and the Economy* 15 (1): 115–152.

———. 2015b. "Skilled immigration and the employment structures of US firms." *Journal of Labor Economics* 33 (S1): S147–S186.

Kerr, William R. 2008. "Ethnic scientific communities and international technology diffusion." *The Review of Economics and Statistics* 90 (3): 518–537.

Kindleberger, Charles P. 1996. *World economic primacy: 1500-1990.* Oxford University Press.

Landes, David S. 1998. *The wealth and poverty of nations: Why are some so rich and others so poor.* New York: WW Norton.

Laudel, Grit. 2003. "Studying the brain drain: Can bibliometric methods help?" *Scientometrics* 57 (2): 215–237.

Maddison, Angus. 2007. *The world economy volume 1: A millennial perspective volume 2: Historical statistics.* Academic Foundation.

Manning, Alan, and Barbara Petrongolo. 2017. "How local are labor markets? Evidence from a spatial job search model." *American Economic Review* 107 (10): 2877–2907.

Manuel, Frank E. 1968. *A portrait of Isaac Newton.* Cambridge MA: Harvard Univ. Press.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics (Chapter 2)*, 65–119. P. Zarembka, Academic Press: New york.

———. 1978. "Modeling the choice of residential location." *Transportation Research Record,* no. 673.

McKenzie, David, and Hillel Rapoport. 2007. "Network effects and the dynamics of migration and inequality: theory and evidence from Mexico." *Journal of Development Economics* 84 (1): 1–24.

Michaud, Joseph-François. 1811. *Biographie universelle ancienne et moderne, 45 vols.* Paris: Bureau de la Biographie Universelle.

Miguelez, Ernest, and Rosina Moreno. 2013. "Do Labour Mobility and Technological Collaborations Foster Geographical Knowledge Diffusion? The Case of European Regions." *Growth and Change* 44 (2): 321–354.

Mokyr, Joel. 2010. *The Enlightened economy: an economic history of Britain 1700-1850.* Yale University Press.

———. 2016. *A culture of growth: the origins of the modern economy.* Princeton University Press.

Monroe, Paul. 2000. *Encyclopaedia of History of Education.* New Delhi: Cosmo.

Nadal, Joseph Cyprien. 1861. *Histoire de l'Université de Valence, et des autres établissements d'instruction de cette ville, etc.* Valence: Impr. E. Marc Aurel.

Navarro-Brotons, Victor. 2006. "The Cultivation of Astronomy in Spanish Universities in the Latter Half of the 16th Century." In *Universities and Science in the Early Modern Period,* 83–98. Springer.

Neal, Derek. 1999. "The complexity of job mobility among young men." *Journal of Labor Economics* 17 (2): 237–261.

Ortega, Francesc, and Giovanni Peri. 2013. "The effect of income and immigration policies on international migration." *Migration Studies* 1 (1): 47–74.

Özak, Ömer. 2010. "The voyage of homo-economicus: Some economic measures of distance." Department of Economics, Brown University.

———. 2018. "Distance to the pre-industrial technological frontier and economic development." *Journal of Economic Growth* 23 (2): 175–221.

Pierson, Alice Sharp, and Peter Cotgreave. 2000. "Citation figures suggest that the UK brain drain is a genuine problem." *Nature* 407 (6800): 13.

Rangeard, Pierre, and Albert Lemarchand. 1868. *Histoire de l'Université d'Angers*. E. Barassé.

Rashdall, Hastings. 1895. *The Universities of Europe in the Middle Ages*. Oxford: Clarendon Press.

Revelt, David, and Kenneth Train. 1998. "Mixed logit with repeated choices: households' choices of appliance efficiency level." *Review of Economics and Statistics* 80 (4): 647–657.

Sarrias, Mauricio, et al. 2016. "Discrete choice models with random parameters in R: The Rchoice Package." *Journal of Statistical Software* 74 (10): 1–31.

Schmutz, Benoit, and Modibo Sidibe. 2019. "Frictional Labour Mobility." *Review of Economic Studies* 86 (4): 1779–1826.

Serafinelli, Michel, and Guido Tabellini. 2017. "Creativity over time and space." mimeo.

Sergescu, Petre. 1939. "Les mathématiques à Paris au moyen-âge." *Bulletin de la Société Mathématique de France* 67:27–42.

Slottved, Ejvind. 1978. *Lærestole og lærere ved Københavns Universitet 1537-1977.* Copenhagen: Samfundet for dansk Genealogi og Personalhistorie.

Sommervogel, Carlos. 1890. *Bibliothèque de la Compagnie de Jésus*. Brussels: Oscar Schepens.

Spitzer, Yannay, and Ariell Zimran. 2018. "Migrant self-selection: Anthropometric evidence from the mass migration of Italians to the United States, 1907–1925." *Journal of Development Economics* 134:226–247.

Squicciarini, Mara P, and Nico Voigtländer. 2015. "Human capital and industrialization: evidence from the age of enlightenment." *Quarterly Journal of Economics* 130 (4): 1825–1883.

Stelter, Robert, David de la Croix, and Mikko Myrskylä. 2021. "Leaders and Laggards in Life Expectancy among European Scholars from the Sixteenth to the Early Twentieth Century." *Demography* 58:111–135.

Stephan, Paula E., and Sharon G. Levin. 2001. "Exceptional contributions to US science by the foreign-born and foreign-educated." *Population Research and Policy Review* 20 (1-2): 59–79.

Sussman, Nathan. 2006. "Income inequality in Paris in the heyday of the commercial revolution." Technical Report, DEGIT (Dynamic Economics, Growth, and International Trade).

Taisand, Pierre. 1721. *Les vies des plus celebres jurisconsultes de toutes les nations, tant anciens que modernes, savoir, latins ou romains, francois, espagnols, italiens, allemans, anglois, hollandois.* Sevestre.

Trippl, Michaela. 2013. "Scientific mobility and knowledge transfer at the interregional and intraregional level." *Regional Studies* 47 (10): 1653–1667.