# Contemporaneous Location Choice of University Scholars in Italy

---


Thesis presented by

**Chiara Zanardello**

Supervisor

**David de la Croix** (UCLouvain/UNamur)

Supervisor and Reader

**Massimiliano Bratti** (UNIMI)


Academic year 2020/2021

In order to obtain the Joint Degree

**Master 120 en Sciences économiques, Orientation générale, Finalité approfondie** (UCLouvain/UNamur)

and

**Dottore magistrale in Economics and Political Science** (UNIMI)

---


Economics School of Louvain/UCLouvain • Place Montesquieu 3 • 1348 Louvain-la-Neuve / Belgium

Département des Sciences économiques/UNamur • Rempart de la Vierge 8 • 5000 Namur / Belgium

Università degli Studi di Milano • Via Festa del perdono 7 • 20122 Milano / Italy

## Abstract

This thesis investigates the contemporaneous academic market in Italy, mapping current scholars' location choice. A unique dataset of contemporaneous professors is built, associating each academic with a composite indicator of their quality. Moreover, the analysis involves universities' quality and the features of the city in which the institution is located. Locations of universities are taken as given in this research. Different patterns are studied: distance, agglomeration (scholars are attracted by universities with higher quality), selection (better scholars travel longer distances), and sorting (the better the scholar, the more the quality of universities is weighted). The findings are confirming literature results and disclosing a new possible tendency for the future academic market in Italy.

## Acknowledgments

From the beginning of this thesis, I could count on great support. First of all, I would like to thank my supervisor, Professor David de la Croix, for the opportunity to work with him. Your feedback has always been precise and pushed me to think further, to try other perspectives without ever stopping at the first sight. A sincere thanks to my reader, Professor Massimiliano Bratti, who provided thoughtful improvements for my research in the most punctual way. Moreover, I would like to thank Professor Daniele Checchi for addressing me in the data collection and Professor Fabrizio Iacone, for the support and the interest showed in this project. My thoughtful thanks are then directed to all professors and researchers who replied to me during the data collection: your affectionate answers enriched a lot my motivation.

My family and friends have had a crucial role in this path, too. The biggest thank is for my parents, Cristina and Francesco: without you, nothing of what I have accomplished so far would have been possible. You were always there, giving me the space to grow and the freedom to choose my own path. Thanks to my brothers as well. Davide, you were the first to push me to go abroad to widen my horizons, and Fabio, thank you for all the solutions we found together. Moreover, also my relatives have been essential in this progress. Studying far away from my town would not be possible without you: Marta and Roberto, with Cate and Vicky, thanks for sharing your home and your time with me. Tina and Gianni, you were already there at the beginning of this journey, I can not thank you enough.

Finally, my thanks are also for those friends who lightened the pressure during these peculiar years. Silvia, Giulia, and Chiara from home. Martina, Michela, and Anna in Milan first and then even from distance. Cecilia, Ines, Joan, Pietro, Silas, Luis, and Andrea met in Louvain-la-Neuve and became my new family. It was a pleasure sharing this journey with you.

# Table of Contents

# 1 Introduction

Knowledge and its mobility have a relevant impact on societies. Understanding the location choice of scholars, as those embodying the upper tail of the human capital distribution, is crucial. In fact, high level of education fosters new developments and fast enhancements (Barra and Zotti, 2018; Nelson and Phelps, 1966; Barro, 2001), improving every-day life (Squicciarini and Voigtländer, 2015; Mokyr et al., 2020).

This research is interested in the location choice of contemporaneous Italian academics as a function of distance and quality, given the geographical position of universities. In addition, once the current academic market is mapped, its peculiarities are compared with those of the past Italian academia.

A Random Utility Model (RUM) is employed. Empirically, through a unique dataset, a multinomial logistic regression allows comparing scholars' utility of living in different regions with respect to their birthplace. In addition to the standard distance effect, appearing to be negative as in the literature (Beine et al., 2011), the research includes also attractive forces linked to the university or to the city in which it is located (size and welfare status). Once fixed effects are excluded, the quality of university (*notability* hereinafter) results to strongly attract Italian scholars (agglomeration effect). Together with the welfare status of the city, these two forces are highlighting the presence of *agglomeration* in the academic world of the peninsula. Nevertheless, this is slowed down by the negative sign associated with the size of the city, which is implying a dispersion effect, although weaker than agglomeration. The attraction of high-skilled people is essential to create a dynamic context and consequent positive spillovers for the society (Kerr et al., 2016; Stephan and Levin, 2001; Grogger and Hanson, 2015; Kerr et al., 2017).

Furthermore, better-educated individuals are also those who detain higher incentives to move, given their higher growth perspectives (Handler, 2018; Barrientos, 2007; Docquier and Marfouk, 2006; Faggian and McCann, 2009; Zhao et al., 2021). Evidence of positive selection and positive sorting are found for the Italian academia, too. In fact, *positive selection*, which is defined by the interaction between distance and individual quality, is solid. It is confirming the fact that the higher is the individual quality (*human capital* hereinafter), the more incentives the scholar has to move, travelling towards better destinations with respect to her original location. Instead, *positive sorting*, described by the interaction between human capital and notability, is weaker than the previous result. This weakness is due to the historical structure of the Italian high-education system, which is still influenced by the traditionally centralized and seniority-based apparatus (Rebora and Turri, 2008; Capano, 2008; MacLeod and Urquiola, 2021). Reforms to increase the autonomy of universities (DPR n. 390/1998) are too recent to be detected in the current project, hence only a low significance level is found when sorting is included in the analysis. This structural feature explains also why, when past scholars (lived between 1734 and 1800) are considered, no sorting effect can be disclosed.

This analysis contributes to the wide migration literature, focused on knowledge-based mobility (Stephan and Levin, 2001; Beine et al., 2011; Grogger and Hanson, 2015; Zhao et al., 2021; MacLeod and Urquiola, 2021). In particular, it aims to integrate the literature about mobility in the Italian academic world. To the best of my knowledge, only student mobility has been analysed (Agasisti and Bianco, 2007; Triventi and Trivellato, 2008; Bratti and Verzillo, 2019), but none has investigated the drivers of scholars' location choice in Italy.

# 2 General framework

## 2.1 Baseline research

This research is an extension to current times of the project on the *"Upper Tail Human Capital and the Rise of the West"*[1]. In the paper titled "The Academic Market and the Rise of Universities in Medieval and Early Modern Europe" (de la Croix et al., 2020), the authors have taken into consideration the period running from 1000 CE until 1800 CE. Through the understanding of that academic market, they underlined the relevance of the "upper tail human capital" [UTHC] in the socio-economic growth experienced by Europe since the Scientific Revolution. The research question of their work is whether the location choice of scholars had significantly influenced these signs of progress and developments.

The unique dataset they have built allowed them to present three major results, describing the mobility of scholars at that time. The first is *agglomeration*, meaning that scholars are attracted by high-quality universities. They found positive and significant coefficients for all the variables included: notability of the institution, economic development of the city measured using the size of its population as a proxy, and level of democracy summarised by communal freedoms. The second result is *positive sorting*, implying that better scholars weigh more the quality of universities compared to those with a lower human capital index. The third outcome is *positive selection*, highlighting how the cost of distance from the birthplace to the university location is weighted less if a high-quality scholar is considered with respect to a low-quality one: better professors are moving over longer distances.

The description just delivered is the framework from which this thesis is going to analyse the contemporaneous characteristics of the academic market in Italy.

[1]Research program funded by an ERC advanced grant from the European Union, ERC grant 883033.

## 2.2 The Italian case

Italy hosts the oldest University in Europe[2] and has a great historical tradition of literates and scholars, who can be included in the upper tail of the human capital distribution. Nevertheless, the Italian academic system had been centralized for a long time, with a strong influence of who was in charge at that moment. This process increased the importance of hierarchy inside the Italian academia, built through informal relationships among the most important chaired scholars and the contemporary ministry (Rebora and Turri, 2008; Capano, 2008). The mentioned centralization of the system was meant to flatten the great inequalities present in the Italian educational system (Triventi and Trivellato, 2008) and some positive results consist in a more favorable social fluidity (Barone and Guetto, 2016) and in a convergence trend of performance among geographical areas (Baldissera and Cornali, 2020). Despite these efforts, the academic world remains seniority-based (Rebora and Turri, 2008), not only in Italy but in the whole Europe (MacLeod and Urquiola, 2021). In 1946, to develop the functionalities of the system and to enhance further equality, the Italian Constitution defined the universities' autonomy principle. This precept aimed to underline local excellence (Checchi and Verzillo, 2014), giving to each university the autonomy to hire an eligible professor. However, this Constitutional principle entered into force only at the end of the 90s, due to the lack of technical standards. The actual implementation of the reforms[3] transformed the Italian academic world into a more fragmented one, with still some signs of the previous seniority-based apparatus (Rebora and Turri, 2008; Bertola and Sestino, 2011; Perotti, 2008; Bini and Chiandotto, 2003). Among the other modifications, it is important to notice that from Berlinguer's decree (DPR n. 390/1998) the recruitment procedure shifted from a national to a local process[4]. Nowadays, each institution has the autonomy to select candidates eligible to be hired (Checchi and Verzillo, 2014; Durante et al., 2011).

In a system where seniority is the main driver of academic career, quality and individual ability may be irrelevant. However, Checchi et al (2014) brought evidence of the opposite. The authors showed how better scholars are also those responding the most to an increase in the level of competition inside the university world, even when weak incentives are provided.

The literature about mobility in the Italian academic world is poor, to the best of my knowledge it is focused on student mobility (Agasisti and Bianco, 2007; Bratti and Verzillo, 2019; Triventi and Trivellato, 2008), but none investigates the driver of professors' location choice.

[2]University of Bologna, founded as a university in 1088.

[3]Among the others, the most important ones are laws n. 168/1989 and n. 390/1998.

[4]For a detailed explanation of the recruitment process in Italy see the following web page of the Ministry of Education, Universities and Research (only in Italian): https://www.miur.gov.it/reclutamento-nelle-universita.

## 3 Research question

The Italian university context has interesting peculiarities, as just mentioned. After having defined the characteristics of scholars' mobility in Italy, another aim of this research is to understand how these specificities have changed with respect to the past, comparing these results with those of the baseline research (appropriately arranged).

To reach this final objective, firstly it is necessary to investigate the actual structure of Italian academia. The role of the location of higher-education institutions (Barra and Zotti, 2017; Audretsch, 1998; Drucker and Goldstein, 2007; Agasisti et al., 2019; Cottini et al., 2019) is considered to be exogenous. Reasonably, the results of the baseline research appear to be applicable in the recent Italian context, too. Hence, agglomeration should still be found. Nevertheless, the distance covered by contemporaneous academics could appear to be shorter than in the past, with a lower magnitude. In fact, the local appointment of professors may have increased the probability to find local excellence (Checchi and Verzillo, 2014), and eventually the importance of networks and nepotism as pointed out by Durante et al., 2011. Still, it is possible to investigate agglomeration forces, such as notability of the university, but also the attractiveness of the city in which the institution is located, measured with the magnitude of the population (istat.it) and the local disposable income of private households (finanze.gov.it). Thus, the following hypothesis is defined:

**Hypothesis 1:** *Agglomeration; scholars are attracted by universities with higher notability.*

Regarding the second result, it is plausible to think that better scholars have also better perspectives of career, and their expected gains are higher in high-quality environments. Thus, in the cost-benefit analysis, the weight assigned to the notability of the university should be higher for better professors[5] than in the analysis of academics with lower human capital. For this, the second hypothesis is delineated as follows:

**Hypothesis 2:** *Positive sorting; scholars with higher human capital weigh more the notability of universities with respect to scholars with lower human capital.*

For what concerns selection, both the literature and the baseline research show how better-educated people are moving more than others (Schiller and Cordes, 2016; Zhao et al., 2021), in turn, better professors may travel longer distances, too. Given the geographical configuration of Italy, it is interesting to investigate whether the quality of individual academics is a driver for their location choice.

[5]Here, the term 'professor' must be intended as a synonym of 'scholar/academic'.

Hence, hypothesis 3 is characterized:

**Hypothesis 3:** *Positive selection, scholars with higher human capital move over longer distances than scholars with lower human capital.*

## 4 Data Sampling

The research is based on a unique dataset. The process of data collection started from RePEc[6] website, using as reference the ranking called "Top 25% Institutions and Economists in Italy". It lists the best universities and institutions among those included in EDIRC (Economics Departments, Institutes, and Research Centers in the World), which considers not only universities as such, but also public agencies, central banks, independent research centres, and associations (for more details see section 2.3 in Zimmermann, 2012). Each institution gains from every author's affiliation RePEc collects, implying an advantage for more populous entities (section 6 in Zimmermann, 2012; Seiler and Wohlrabe, 2011).

**Universities.** For the present work, only universities will be taken into account[7]. In particular, those included in the research are: University of Bologna (UNIBO), Catholic University of the Sacred Heart (CATT), University of Verona (UNIVR), University of Catania (UNICT), University of Milan (UNIMI), University of Rome - Tor Vergata (UNIROMA2), University of Florence (UNIFI), University of Venice (UNIVE), Polytechnic University of the Marches (UNIVPM), Sapienza University of Rome (UNIROMA1), University of Turin (UNITO), University of Trento (UNITN), University of Naples Federico II (UNINA), University of Padua (UNIPD), Bocconi University (BOCCONI), University of Genoa (UNIGE), University of Palermo (UNIPA), Free University of Bozen (FUB), University of Bari (UNIBA).
Among these, 16 are public universities, 1 is a polytechnic (UNIVPM), and 2 are privately founded (BOCCONI and CATT).

**Scholars.** Each institution includes a list of members (registered in the RePEc Author Service) and these are the observations considered when building the dataset[8]. The people registered to the server have different roles inside the academia. For the current research, only professors and research fellows (also postdoctoral) are taken into account[9]. As for the former, full and associate professors are considered, but also adjunct and assistant ones. Moreover, there are also few emeritus professors, who are still teaching. In fact, the *"teaching disclaimer"* is used

[6]Research Papers in Economics (RePEc) is a project collecting bibliographic data about papers in economics and similar fields, aiming to spread and enhance relative researches.
[7]From here onwards the words 'university' and 'institution' are considered as synonyms.
[8]The ranking is updated month by month; hence the names collected (and the status associated to them) can be changed with respect to the period of data collection, which is approximately December 2020 – April 2021.
[9]After having determined the status of each member, doctoral students are excluded from the dataset.

to handle the exceptions. This means that whenever a particular case occurred, the scholar was considered in the sample only if she is involved in teaching activities. This circumstance refers to the aforementioned cases of emeritus professors, but as well to academics taking part in visiting programs or national/international collaborations. Hence, for example, a visiting professor is considered as a scholar also in the foreign university only when she explicitly mentions her teaching activity there, too (more on multiple affiliations later). Moreover, cases of scholars "on leave" were not considered part of the sample, given the absence of the *teaching disclaimer*. This way to proceed left out the consideration of research centres, like CEPR, IZA, CESifo, assuming the honorific character of their appointments. Table 1 presents the precise taxonomy for the scholars included in the dataset, with the respective quantities and percentages.

Table 1: Taxonomy of scholars

| **Categories** | **Quantity** | **Percentage** |
|---|---|---|
| Full professors | 386 | 38.79% |
| Associate professors | 281 | 28.24% |
| Assistant professors | 98 | 9.85% |
| Adjunct professors | 43 | 4.32% |
| Research fellows | 108 | 10.85% |
| Post-doctoral fellows | 26 | 2.61% |
| Emeritus professors | 8 | 0.80% |
| Visiting professors | 45 | 4.52% |

**Affiliations.** Once listed, the observations were associated with their university. The association process required a careful investigation for each academic. The Curriculum Vitae (CV) was the main source exploited, however not always it has been sufficiently updated or complete. Thus, also LinkedIn[10] was consulted, in addition to personal web pages (institutional and/or private), when available. Once the sources were found, the association was collected considering the most updated affiliation at the moment of consultation[11]. Affiliations to telematic universities were not taken into account and, as said, neither to research centres. Only the European University Institute (EUI)[12] and the University School for Advanced Studies in Pavia (IUSSPAVIA)[13] have been considered because they can be equated to actual universities.

For those universities with multiple locations, the main one was considered, assuming that the majority of the scholars teaching in one location are also teaching in the other(s). This can generate some bias when locations are far away from each other; as in the case of Catholic University with four locations, respectively in Milan (main building), Brescia, Piacenza, and Rome; proper robustness check will be implemented (more on this later).

[10]Professional social network: linkedin.com
[11]Consultation period: December 2020 - April 2021.
[12]https://www.eui.eu/en/home
[13]http://www.iusspavia.it/en/web/guest/university-school-for-advanced-studies

**Multiple affiliations.** Some scholars are associated with more than one university, in Italy or abroad. This fact had risen the issue of multiple affiliations, which at the moment are 7.44% of the sample; with four maximum affiliations. In the past (i.e. baseline research), those academics linked with multiple institutions were associated with high-quality scores (de la Croix et al., 2020), whereas nowadays it is more frequent to encounter multiple affiliated scholars with low bibliometric indicators. Usually, these academics are younger and have a postdoc position in a university while teaching in another institution.

Empirically, each affiliation of the same scholar is treated as it was chosen by different individuals, leading to their overestimation with respect to unique-affiliated scholars (see section 6.4). In the following part of the paper, the formers will be called repeated movers (RM), while the latters single movers (SM).

Considering multiple affiliations in this way, the initial sample counts 1347 observations. A cleaning process was needed to eliminate scholars not members of the Italian academia anymore, as well as Ph.D. students and non-teaching emeritus and visiting professors, in addition to on leave ones[14]. At the end of the cleaning process, the sample counts 995 names.

**Dataset.** With this procedure 76 universities were found, among these 39 are foreign universities while the other 37 are Italian institutions. From this set, universities with less than five scholars have been excluded from the sample (55), given their minor relevance for academics' choice. The resulting list is the set of choices each professor of the sample faces when maximising her location decision, which now counts for 21 universities, all of which are Italian. Consequently, the number of scholars in the database decreased to 921 observations, the percentage of multiple affiliations is now only 3.04% and the maximum number of associations decreased to three. From here onwards, this subset is considered in the analysis. Table 2 summarises the differences between the original dataset and the subset obtained after dropping universities with less than five scholars.

Table 2: Comparison between datasets

| | **Original** | **Subset** |
|---|---|---|
| Tot. observations | 1347 | 921 |
| N. Universities | 76 | 21 |
| Obs. after cleaning process | 995 | 921 |
| Obs. with known birthplaces | 890 (89.45%) | 829 (90.01%) |
| Obs. with known education | 964 (96.88%) | 891 (96.74%) |
| Obs. with not-known education | 31 | 30 |
| Multiple affiliations | 7.44% | 3.04% |
| Max n. affiliations | 4 | 3 |

[14]One of the main difficulties met has been understanding the meaning of the various roles and titles indicated by each scholar. The final dataset was built to the best of available knowledge, however, minor errors may still be present.

# 5 Methodology

The research aims to study the features of the academic market in Italy, focusing on scholars' location choices and explaining how various determinants may influence this decision. As in the baseline research, this is done employing a Random Utility Model (RUM), a gravity model widely used in migration analysis (Grogger and Hanson, 2016; Ortega and Peri, 2013). It determines the individual utility of living in a certain region, comparing it to the expected utility from moving to alternative locations (Ramos, 2016). Empirically, a multinomial logit model is applied (Akcigit et al., 2016; Ortega and Peri, 2013).

This section will be divided into three sub-sections to better identify the different issues analysed. As first, *location* will be investigated, with research-specific features and sample characteristics. Secondly, attention is given to *quality analyses*, both at the individual and at the aggregate level. The specific *model* is then defined in the last subsection.

## 5.1 Location analysis

The main part of the analysis requires the investigation of the distance a scholar is willing to travel to reach a certain university to develop her career. Here, the distance is considered as an increasing cost for the individual. The further away she is from her usual context, the longer the distance is, and the higher the cost (Schwartz, 1976).

The starting point is collecting birthplaces of each observation, considered as an observable proxy of scholars' usual life context. Other variables are not observable. For instance, the place where academics' families are living is non-observable information, as well as the location where the partner is working, even if these may be relevant drivers of professors' decisions. Other sources should be used to detect these determinants (i.e. surveys), but their implementation would have required more substantial means than those available at the moment of writing.

As for affiliations, CVs and personal webpages were the main sources, given that neither LinkedIn nor RePEc are providing birthplace information[15]. Nevertheless, professors indicating their place of birth somewhere resulted to be only a little portion of the sample (around 30%). Dealing with alive personalities is different from collecting information about people who lived centuries ago. Nowadays, there are more multimedia sources to be exploited, but also bureaucratic and privacy authorizations, which are essential to protect personal information but also slow down the data collection process. Thus, it has been decided to send an email requesting it, this increased by about 55% the birthplaces collected[16]. Now the birthplace is known for

[15] I would to thanks the RePEc administration for the prompt answers.

[16] I should like to express my heartfelt thanks to those who answered in a so interested manner.

90,01% of the academics (829 observations).

Despite this, the collection of these data resulted to be a clear obstacle since the beginning of the project. For this reason, it is included in the dataset also the location of the institution where each scholar obtained her lowest degree of education available in her files. The information should be considered as a proxy of the birthplace, given that most likely the two coincides or are close enough. This procedure increased the dataset to 891 observations, reaching coverage of 96.74%. For the majority of the sample, the lowest level of education is the bachelor's degree, but some academics mentioned the high school, too. Only for a few observations, the lowest education level available was the master's degree, while for five scholars only information about the Ph.D. is known. For this latter case, it seems not necessary an ad-hoc robustness check, given their small number (only 5 out of 921 observations), and given the location of their Ph.Ds: four of them attended the doctorate in the same university (or in a close one) in which they are teaching, while only one obtained the title abroad. Nevertheless, educational information is present almost always in CVs or LinkedIn profiles, when this information was not public a specific email was sent[17]. However, 30 observations remain without education information (3.26% of the sample) and they will be excluded from the relative sample. Two different regressions are implemented: one investigates the birthplaces and another the locations of the lowest level of education (see section 6).

Once location data are collected, the decimal coordinates are matched[18]. Now the dataset has both $i$ observations associated with a geo-localized birth- and/or education-place, and $k$ geo-localized universities.

### 5.2 Quality analysis

Another important step of the project deals with quality indicators: *aggregate* quality scores (notability) and *individual* bibliometric indexes (human capital) are both collected. The formers are at the university level, while the latters are associated with each scholar. When considering variables as quality, the analysis must follow some assumptions, given that there is not an absolute measure. Outcomes are influenced by the parameters included in each indicator. Hence, firstly it is necessary to describe each bibliometric measure collected, and only later the approach followed in the paper.

**Aggregate quality.** Regarding notability indicators, they may suffer from endogeneity problems. Indeed, university scores are related to the quality (and quantity) of affiliated scholars. The problem is tackled by looking for past indicators. Considering that the average age

[17]Usually the case of emeritus professors, or persons not in the Italian academia anymore. Thus, this procedure helped also to determine the status of these exceptions.

[18]The websites used are: latitudelongitude.org, tuttitalia.it for Italy, while latlong.net for foreign cities.

for Italian academics is 48 years (Elaborazioni su banche dati MIUR, DGCASIS – Ufficio VI Gestione patrimonio informativo e statistica, Morana, 2020) and that usually, the beginning of the career is around 30 years, it has been decided to look for quality indicators of around 20 years ago. In RePEc historic archives, it was possible to find aggregate quality scores for top institutions divided by countries going back to 2007, at maximum. Before that date, only simple/ordinal rankings are available, without any score associated with institutions. Hence, in this paper, it was decided to consider scores of 10 years ago, as of December 2010[19,20]. As mentioned, scores for top universities were collected from per-country rankings. The institutions' country was considered to select the ranking and then the associated score was taken. The scores in these rankings are weighted averages of the credit brought by each affiliated scholar, who decides the percentage applied to each institution she is associated with (for the specific formula see section 6, Zimmermann, 2012). This can generate some biases, for example decreasing the relevance of the main affiliation with the addition of more associations, as pointed out by Seiler and Wohlrabe (2011). Among the 21 universities, it was possible to associate 18 universities with their score, while the other 3 were not included in per country rankings of top institutions. The problem of missing data is handled by assuming the worst score of the ranking (i.e. 93.36). Since RePEc is using reversed indexes, in the sense that the lower is the score, the higher is the quality associated, the current analysis will convert them to have a direct relation between indexes and quality, thus facilitating the interpretation. The notability ($ln_Q$) linked to each university ($k \in K$) can be visualized in Figure 1.

**Individual quality.** Regarding individual bibliometric indicators, nowadays the possible choices are many. Firstly, RePEc makes available the previously mentioned top authors per-country ranking (i.e. "Top 25% Institutions and Economists in Italy"). RePEc computes these human capital scores as the harmonic mean of various rankings based on different factors (section 5, Zimmermann, 2012) and it ranks more than 800 scholars. For the present research, the one as December 2020 is considered[21] (see below for missing data).

In the literature, the traditional way to assess academic quality has been based on indicators provided by *Web of Science* (*WoS* – with its three subject-specific ISI citation databases; Yang and Meho, 2006). For a long time, it was one of the few multidisciplinary databases assigning authors' scores based on citations from an original set of sources (Neuhaus and Hans-Dieter, 2008; Jacso, 2005). The main issue has to do with its relative coverage, indeed only a fraction of sources, although a significative one (i.e. journal literature), is considered (Norris and Oppenheim, 2007). However, for economics and social science, this literature is not the

[19] The research started in December 2020.
[20] The mentioned ranking is available here: https://ideas.repec.org/top/old/1012/
[21] Given that the ranking is updated every month, the current online score could present some differences.

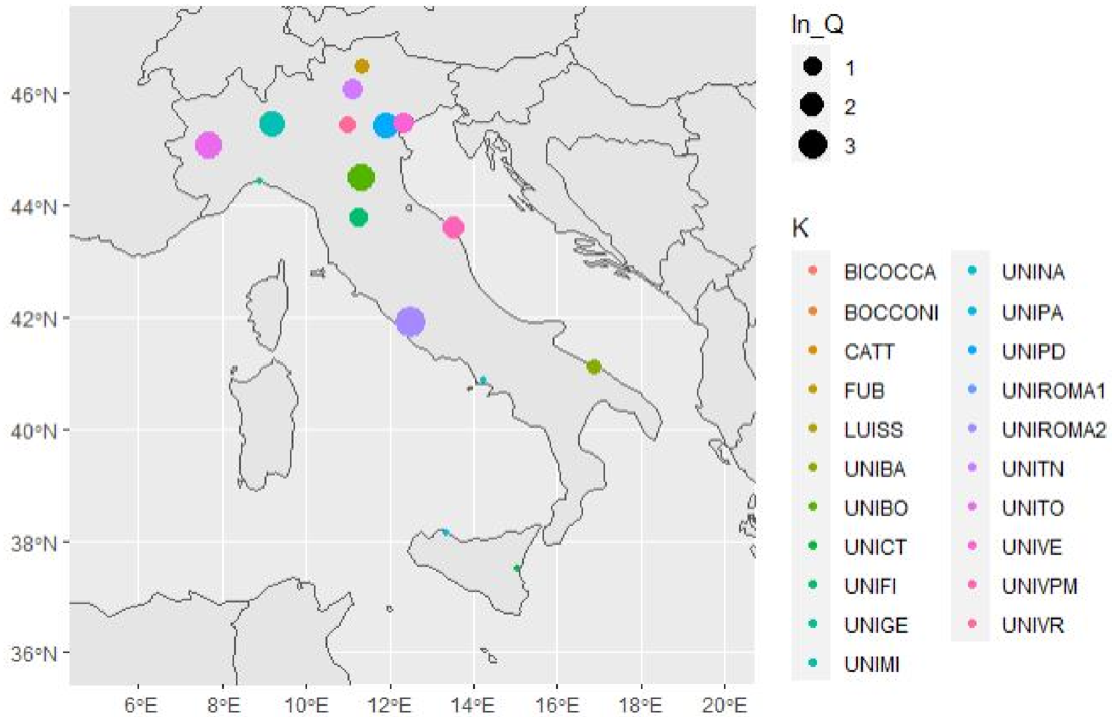


Figure 1: Bubble Plot on Italy map. Showing notability indexes associated to each university: the higher the $ln_Q$, the bigger the bubble on the map. Each point represents the location of $k \in K$ institutions.
Note: BICOCCA, BOCCONI and CATT are overwritten by UNIMI, which has the highest notability in Milan. UNIROMA1 and LUISS are overwritten by UNIROMA2, which has the highest notability in Rome.

main communication driver (Neuhaus and Hans-Dieter, 2008). The social science indicator of *WoS* goes back to 1956[22].

With the automatization of data collection, quality-evaluation possibilities are now augmented with *Scopus*, a database provided by Elsevier, and *Google Scholar*. The former covers a wider range of sources with respect to *WoS*, starting from a predefined database of Elsevier and it goes back to 1996 for social science[23] (Jacso, 2005; Yang and Meho, 2006; Norris and Oppenheim, 2007). The second indicator is a free service by Google, the sources utilized are wide, but not clearly identified by the provider, leading to an important lack of reliability, which is added to weak, imprecise performance, as pointed out by Neuhaus and Daniel (2008). However, being free and having one of the widest coverages among bibliographic indicators, Google scholar has still some potentiality (Neuhaus and Hans-Dieter, 2008).

In addition to these well-known multidisciplinary databases, it has been decided to compare also *WorldCat identities* indexes. Among other information, it stores measures for works (*Worldcat Works*) and library holdings (*Worldcat Library*) for each scholar (and organizations) present in WorldCat.org and OCLC sources (OCLC Research, WorldCat identities[24]).

Every indicator mentioned has its own list of pros and cons, hence a composite indica-

[22] *"Coverage in Web of Science goes back to 1945 for Science Citation Index, 1956 for Social Sciences Citation Index, and 1975 for Arts & Humanities Citation Index."* (Yang and Meho, 2006)
[23] Scopus goes back at maximum to 1966. (Yang and Meho, 2006)
[24] https://www.oclc.org/research/areas/data-science/identities.html

tor is created by combining all of them (RePEc score, Worldcat works and library holdings[25], Google Scholar citations, H-index and i10-index[26], WoS H-index[27], Scopus H-index[28]). The underlined idea is to compare the various human capital scores, understanding the amount of information added by each indicator to the analysis. This is done following a *Principal Component Analysis* (PCA) which aims to reduce the number of variables without losing much accuracy and information. Once the correlation between the variables is computed (Figure 2) and their standardization is completed (remember that Repec indicator must be reversed), the PCA compresses most of the information among the first principal components, which are new uncorrelated variables. For this research, the first two components are taken into consideration, given their standard deviation greater than one. Moreover, the cumulative information explained by these two components is 77.83% of the total (Table 3), claimed to be sufficient. Hence, considering the two components together, the analysis is gaining simplicity while losing only a little portion of its accuracy; from here onwards they are used to represent the new individual quality index.

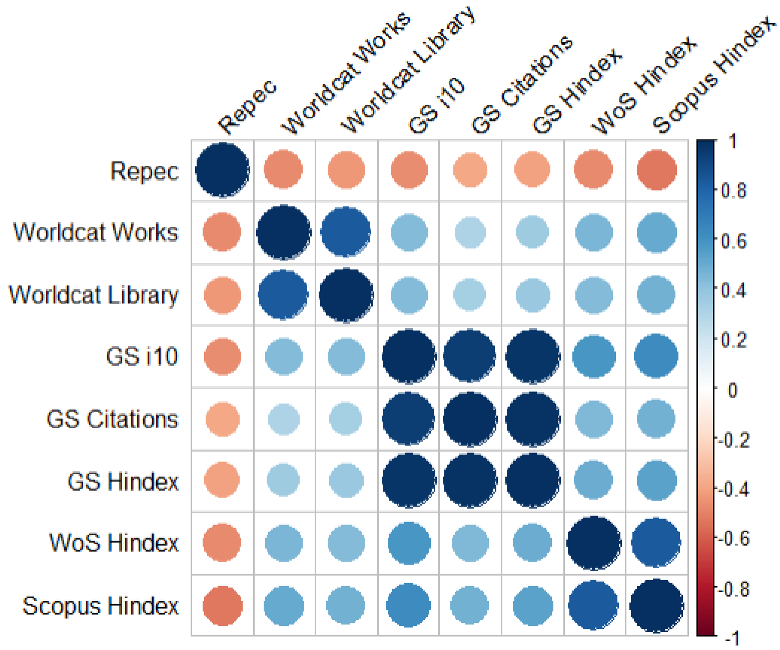


Figure 2: Correlation Matrix Plot. Showing the correlations between the eight different bibiliomeric indicators included in the analysis.

Table 3: Principal Components table. Showing the standard deviation (St.dv.), the proportion of variance (Pr.Var.) and the cumulative proportion (Cum.Pr.) for each principal component.

| | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** | **PC6** | **PC7** | **PC8** |
|---|---|---|---|---|---|---|---|---|
| St.dv. | 2.1920 | 1.1921 | 0.9122 | 0.76239 | 0.40517 | 0.39254 | 0.18920 | 0.08072 |
| Pr.Var. | 0.6006 | 0.1776 | 0.1040 | 0.07265 | 0.02052 | 0.01926 | 0.00447 | 0.00081 |
| Cum.Pr. | 0.6006 | 0.7783 | 0.8823 | 0.95493 | 0.97545 | 0.99471 | 0.99919 | 1.00000 |

[25] https://www.worldcat.org/identities/
[26] https://scholar.google.com/
[27] https://app-webofknowledge-com.pros.lib.unimi.it/author/search?lang=en_US&SID=C45nUucTHDOd2z2VUPs
[28] https://www.scopus.com/freelookup/form/author.uri?zone=TopNavBar&origin=NO\%20ORIGIN\%20DEFINED

### 5.3 The model

At the empirical level, a standard multinomial logit model is implemented. This framework requires perfect elasticity of the demand in the academic market, implying that there is always an available position for every scholar. In the case of Italian academia, this assumption can be interpreted in this way: when a candidate is appointed as eligible to be hired by a university (with Berlinguer's reform - DPR n. 390/1998 - each university has the autonomy to do so), she will be able to find a chair. In fact, local recruitment simplified bureaucratic processes favouring vacancies and the number of available positions increased for each local selection (Checchi and Verzillo, 2014). This assumption can be improved in further developments of the project, given that who is eligible to be hired by a university is who succeeds in the respective local public exam. This means that only scholars with higher individual quality indexes are able to freely choose the location of their career. In the current thesis, the relevance of this fact is reduced by including individual human capital scores in the analysis, which remains a partial equilibrium analysis. Hence, competition and ability factors are considered as demand-side factors. Competition variables are associated with the notability of universities and the features of the city in which it locates: the higher the score, the fiercer the competition among academics. Then, the ability variable is represented by the individual human capital.

A multinomial logit model allows us to compute the probability that a $k$ university, belonging to $K$ set of choices, is maximising an $i$ scholar's utility (McFadden, 1974). The first step consists of defining the utility function for each $i$ scholar. It is defined by a deterministic component $V_{ik} = \beta x_{ik}$, capturing the average of benefits and costs of each location choice, and by a random component $\epsilon_{ik}$, describing unobservable factors that may influence the utility, assumed to be orthogonal to $\beta x_{ik}$. Hence, the utility function can be written as follow:

$$U_{ik} = V_{ik} + \epsilon_{ik} = \beta x_{ik} + \epsilon_{ik} \tag{1}$$

Then, the model requires the independence of career choices, fulfilled with the assumption of $\epsilon_{ik}$ being independent and identically distributed. Now, it is possible to define the main equation of the model, where the probability of a specific choice is confronted with the whole set of available choices, implying that any change in the features of the latter will influence the resulted probability.

$$p_{ik} \equiv [U_{ik} = \max_{k' \in K} U_{ik'}] = \frac{exp(\beta x_{ik})}{\sum_{k' \in K} exp(\beta x_{ik'})} \tag{2}$$

Another important assumption of the model is the Independence of Irrelevant Alternatives (IIA) among the set of choice $K$, thus the choice between two specific alternatives should depend only on their own features, without any external influence of a third one (McFadden, 1974). This means that scholars' choice is depending only on the two considered universities and not on the rest of the university set. Additionally, this assumption is limiting the analysis to choices made within the academic world. This conclusion is plausible, given the impossibility of considering the choice faced by academics when deciding whether to become a professor or to follow other career paths. Notwithstanding, this assumption can not always be justified and in section 6.4 it will be relaxed.

The next step is to define the explicit form of the deterministic component, the one capturing the difference between average benefits and costs of choosing $k \in K$. Regarding the benefits, they are an increasing function of the university's notability $Q_k$ (as defined in the previous subsection), however, also the attractiveness of the city should be considered. Hence, variables such as $P_k$ and $Y_k$, representing respectively cities' total population (capturing the size of the city) and households' disposable income[29] (capturing the welfare status) of the city in which $k$ university is located. In addition, a better scholar (with a high individual quality index $q_i$) has more to gain from a welcoming environment (i.e. a university with a certain $Q_k$), and this is captured by the interaction term between the two variables. Therefore, the benefits equation is:

$$B_{ik} = a_0 + a_1 Q_k + a_2 P_k + a_3 Y_k + a_4 q_i Q_k \tag{3}$$

where $\forall a \in \{a_1, a_2, a_3, a_4\}$ greater than zero.

About the cost analysis, it is important to underline that these are competition costs. Thus, it is evident (considering the distance as a cost for academics, in terms also of family attachment) that the higher the distance from the birthplace (or education-place) the higher is the burden to travel until there, and hence minor will be the competition among scholars. However, the better is an academic (i.e. the higher $q_i$), the higher the gain she may obtain in a certain university environment ($Q_k$), implying a reduction in competition costs. In addition, as shown by the literature, a better scholar should be willing to move longer distances, thus the interaction term between distance and individual quality may increase the cost. Then, the higher is the attractiveness of a city ($P_k$ and $Y_k$) and a university ($Q_k$), the higher will be the

[29]Italy (2018, average disposable income IRPEF, city-level): https://www1.finanze.gov.it/finanze3/analisi_stat/index.php?search_class[0]=cCOMUNE&opendata=yes#

competition. Therefore, the costs equation is:

$$C_{ik} = b_0 + b_1 Q_k + b_2 P_k + b_3 Y_k - b_4 q_i Q_k - b_5 d_{ik} + b_6 d_{ik} q_i \tag{4}$$

where $\forall b \in \{b_1, b_2, b_3, b_4, b_5, b_6\}$ greater than zero.

The equations for the benefits and the costs are defined, hence, it is possible to give the explicit form of the deterministic component, specifying for each dyadic match (i.e. the association of scholar $i$ with university $k$) the net benefit. This is done by subtracting (4) to (3), with the addition of a fixed effect $\gamma_k$. The subscript $k$ suggests that it is referring to non-measurable universities' (cities') characteristics which may influence their ability to attract and are not varying across time[30]. These fixed effects are perfectly identifying agglomeration variables ($Q_k$, $P_k$, $Y_k$) included in the model, given their similar characteristics (destination-specific and time invariant). For this reason, fixed effects are excluded when agglomeration forces are considered (more on this later). Hence, the general expression obtained is:

$$\beta x_{ik} \equiv V_{ik} \equiv B_{ik} - C_{ik} = \beta_0 + \beta_1 Q_k + \beta_2 P_k + \beta_3 Y_k + \beta_4 q_i Q_k + \beta_5 d_{ik} + \beta_6 d_{ik} q_i + \gamma_k \tag{5}$$

$\beta$ is a vector, whose parameters are those of interest and will be estimated. In particular, the constant in equation (5) can be specified as the difference between the two constants in (3) and (4). Then, to define *agglomeration effect*, it is relevant to look at the notability of universities and the attractiveness of cities, represented by $\beta_1, \beta_2, \beta_3$ computed as $a_j - b_j$, with $j = \{1, 2, 3\}$. If the sign found is positive then agglomeration is characterized, instead, if it is negative there may be evidence of dispersion. *Sorting effect* is measured by $\beta_4 \equiv a_4 + b_4$. If it is positive, it means that higher individual quality reduces the cost or increases the gain to travel towards better universities. $\beta_5 \equiv -b_5$, is the coefficient capturing the expected effect of distance, considered a cost, as previously mentioned. $\beta_6 \equiv b_6$ underlines *selection effect*, when there is a positive sign better scholars are moving along further distances with respect to other academics.

Equation (5) includes only destination-specific regressors, linked directly or interacted with universities' features. In fact, human capital ($q_i$) alone is never present, because it would not be identified, influencing all dyadic matches in a symmetric manner.

[30]Fixed effects are used to overcome omitted variables biases, they are controlling for unobserved variables which are not changing over time. If there is a change over time, they can be inefficient, with large standard errors. (Williams, 2019)

## 6 Analytical results

### 6.1 Main regressions

The multinomial logit model described above is used to estimate the main regressions of the research. It includes scholars for whom the place of birth is known (829 observations - 90.01% of the sample), or academics associated with the location of their lowest level of education (891 observations - 96.71% of the sample). Each location is geo-localized, and each academic is matched with a unique individual quality index, computed with a PCA (see section 5.2). Furthermore, it must be remembered that universities with less than five professors have been deleted from the database, assuming their minor relevance in the total set of choices. Now, the latter is counting for 21 geo-localized institutions, which are linked to their RePEc quality score (see section 5.2). The estimation does not allow for zero indexes both at aggregate and at individual level (logarithms are taken). Thus, if a scholar is not matched with any positive score, she is associated with the worst human capital index of the sample (794,82 for RePEc, 1 for all the others indicators). In fact, it is reasonable to assume that she is not publishing as much as those colleagues with a positive score. Nevertheless, it could also be the case that the sources collected to compute bibliometric indicators are not capturing her work, a weakness known in quality analyses. The same is applied to universities with indexes at zero, which are linked with the worst positive score of notability, and in this case the index is reversed (see section 5.2). The transformation in logs is applied to distance as well. Hence, there is the issue of zero distances, related to those scholars born in the same city in which they are teaching. These academics bear the minimum cost of distance, which is assumed to be the same as in the baseline research: 3,5 km - walking distance from the Vatican city to the Colosseum, in the old city of Rome (de la Croix et al., 2020).

In the two following subsections, the results of the main regressions are described, considering first the birthplaces and then the locations of the lowest level of education. The package used is called "mlogit" by Croissant (2020). The evaluation is focused on the sign and on significance of the coefficients indicating distance, agglomeration, selection, and sorting effects. Fixed effects are controlled in each regression, apart from when the agglomeration effect is introduced. In this latter case, two variables, capturing the characteristics of the city in which the university is located ($P_k$ and $Y_k$ - see section 5.3) and one representing the reputation of the university itself ($Q_k$), are included. Some descriptive statistics for both analyses is presented in Table 4.

Table 4: Descriptive statistics

| Variables | Obs* | Mean | Median | St.Dv. | Variance |
|---|---|---|---|---|---|
| Birthplace analysis: | | | | | |
| ln of distance | 17409 | 5.621046 | 5.869873 | 1.259713 | 1.586877 |
| ln of human capital | 17409 | 0.14579 | 0.01207 | 2.14257 | 4.59062 |
| ln of notability | 17409 | 1.26286 | 1.11871 | 0.96520 | 0.93161 |
| ln of population | 17409 | 13.04431 | 13.25533 | 1.55534 | 2.41907 |
| ln of income | 17409 | 10.11043 | 10.10835 | 0.16147 | 0.02607 |
| Lowest level of education analysis: | | | | | |
| ln of distance | 18711 | 5.45385 | 5.71821 | 1.48628 | 2.20903 |
| ln of human capital | 18711 | 0.02782 | -0.1225 | 2.18878 | 4.79076 |
| ln of notability | 18711 | 1.26286 | 1.11871 | 0.96520 | 0.93161 |
| ln of population | 18711 | 13.04431 | 13.25533 | 1.55534 | 2.41978 |
| ln of income | 18711 | 10.11043 | 10.10835 | 0.16147 | 0.02607 |

Note: *Obs represent the number of possible dyadic matches in each analysis.

### 6.2 Main results - Birthplaces analysis

Table 5 shows the results of the multinomial logit estimations regarding birthplaces analysis. The dataset counts 829 scholars (90.01% of the sample) choosing among 21 universities, resulting in 17409 possible dyadic matches.

The first column contains the basic equation of gravity, highlighting the negative sign of *distance* coefficients, $ln_d$. This implies that the greater is the distance between the birthplace and the location of the university, the higher are the competition costs and the lower is the probability of finding a dyadic match. The coefficients remain highly significant in every specification. The magnitude is coherent with the one found in contemporary migration literature: distance coefficients around 0.7 can be found in *"Diasporas"* (by Beine et al., 2011), when the migrants are not divided between low- and high-skilled ones. However, this coefficient is lower with respect to those analysing past periods.

*Selection* effect is added in the second column. This is defined by the interaction term between human capital and distance, $ln_q ln_d$. The sign is positive as expected, meaning that scholars with the higher human capital move over longer distances with respect to scholars with lower human capital. The high significance of the coefficient (at 1%) is confirming the third hypothesis of *positive selection*, which is found in each specification of the model.

Column (3) is showing the effect of *sorting*, through the interaction between individual human capital and university notability ($ln_q ln_Q$). The positivity of the coefficient shows evidence for *positive sorting*, as expected in the second hypothesis. Despite this, the significance of sorting is only at 10%: sorting appears to be weaker than selection. However, both coefficients maintain their significance level also in column (4), which considers both effects simultaneously. Then, comparing log-likelihood (LL) values allows to compute a likelihood-ratio (LR) test: considering column (4) over column (1) the null hypothesis of no selection and no sorting

Table 5: Multinomial logit regressions: standard logit model - birthplaces analysis

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance:** | | | | | | |
| $ln_d$ | $-0.780^{***}$ | $-0.795^{***}$ | $-0.781^{***}$ | $-0.794^{***}$ | $-0.773^{***}$ | $-0.787^{***}$ |
| | (0.027) | (0.027) | (0.027) | (0.027) | (0.026) | (0.027) |
| **Selection:** | | | | | | |
| $ln_q ln_d$ | | $0.048^{***}$ | | $0.047^{***}$ | | $0.047^{***}$ |
| | | (0.012) | | (0.012) | | (0.012) |
| **Sorting:** | | | | | | |
| $ln_q ln_Q$ | | | $0.036^{*}$ | $0.031^{*}$ | | $0.035^{*}$ |
| | | | (0.018) | (0.018) | | (0.019) |
| **Agglomeration:** | | | | | | |
| $ln_P k$ | | | | | $-0.174^{***}$ | $-0.178^{***}$ |
| | | | | | (0.026) | (0.026) |
| $ln_Y k$ | | | | | $2.471^{***}$ | $2.567^{***}$ |
| | | | | | (0.321) | (0.323) |
| $ln_Q$ | | | | | $0.232^{***}$ | $0.228^{***}$ |
| | | | | | (0.042) | (0.042) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs | 829 | 829 | 829 | 829 | 829 | 829 |
| $R^2$ | 0.210 | 0.214 | 0.211 | 0.215 | | |
| LL | −1,844.042 | −1,835.291 | −1,842.127 | −1,833.845 | −1,952.814 | −1,942.488 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

is rejected.

To investigate *agglomeration*, university fixed effects must be excluded from the regression (columns (5)), because otherwise, without time variation, they perfectly identify agglomeration variables. Moreover, without fixed effects, it is possible to study the relevance of the attractiveness of cities in which universities are located. All the three variables included are highly significant. The logarithm of the populations ($ln_P k$) is negative, preluding a presence of dispersion: the probability that a scholar chooses university $k$ decreases with the increase of the city size. The logarithm of disposable income ($ln_Y k$) is instead positive, implying a great attractive force of the variable: the richer the city, the greater is the likelihood a professor develops her career in that institution. The logarithm of university notability ($ln_Q$) is also significant at 1% and positive, meaning that the higher is the reputation of the university the higher is the possibility that a scholar will move there. Therefore, the first hypothesis about agglomeration holds: even if from $ln_P k$ there is a tendency for dispersion (given its negative sign), this is greatly compensated by the attractive force of the other two variables involved.

Signs, significance levels, and magnitudes characterizing each effect are maintained also in the sixth column, where all coefficients are considered.

### 6.3 Main results - Lowest level of education analysis

The results of multinomial logit estimations considering the location of the lowest level of scholars' education are presented in Table 8 (in the Appendix). Associating 891 observations with 21 universities, now the dataset counts 18711 dyadic matches.

The only coefficients that remain significant at 1% are those referred to *distance* and *agglomeration*. The sign of the former is still negative and the magnitude of about 0.7 is confirmed in each specification, although it appears to be slightly decreased compared to the birthplaces analysis. From the models without fixed effects (columns (5) and (6)), all the agglomeration variables ($ln_P k$, $ln_Y k$, $ln_Q$) again confirm the first hypothesis, with the same signs as in the birthplaces analysis.

For what concerns *selection effect* ($ln_q ln_d$), its coefficients are still positive, but not significant anymore. Similar evidence is found for *sorting* ($ln_q ln_Q$), which presents positive signs, but not significant coefficients. Hence, these results are proving that the second and the third hypothesis are confirmed only if the actual location of birth is taken into account; although the LR test between (4) and (1) is still rejecting the null hypothesis of no effects. On the other hand, for the standard effect of distance, the lowest level of education can be considered as a proxy of the birthplace. This reasoning holds also for agglomeration effects, given that in this case the analysis is focused on features of universities (reputation/quality) and of the cities in which they are located (population size and income level), both aspects do not vary much with respect to the previous analysis. Selection and sorting are both interacted with the individual level of quality, which is affected more by the change of dataset.

Given the presented results for both analyses, the one involving scholars' birthplaces is considered to be more relevant for the project. Thus, only that one will be taken into account as the benchmark model in the following part of the paper, where further robustness checks are developed.

### 6.4 Robustness checks

**Age.** In the benchmark model, the age of the scholars is not considered. This might be an important weakness of the previous estimation. Hence, it has been decided to include in the dataset also professors' age. This meant an additional search on CVs, LinkedIn, personal/institutional web pages. CVs resulted to be the main source, in fact, most of the scholars indicated the exact year of birth among their personal information. However, 30.94% of the dataset remains without age information. For these cases, the final year of the Ph.D. is collected, assuming that scholars at the end of their doctorate are 30 years old. Once all sources were exploited, 28 observations did not have any age reference, neither the exact year of birth nor the last year of Ph.D.. These cases are excluded from the following analysis, bringing the number of dyadic matches at 16821, corresponding to 801 academics. The average age is 50 years and few months, Figure 3 is presenting the age distribution of scholars.

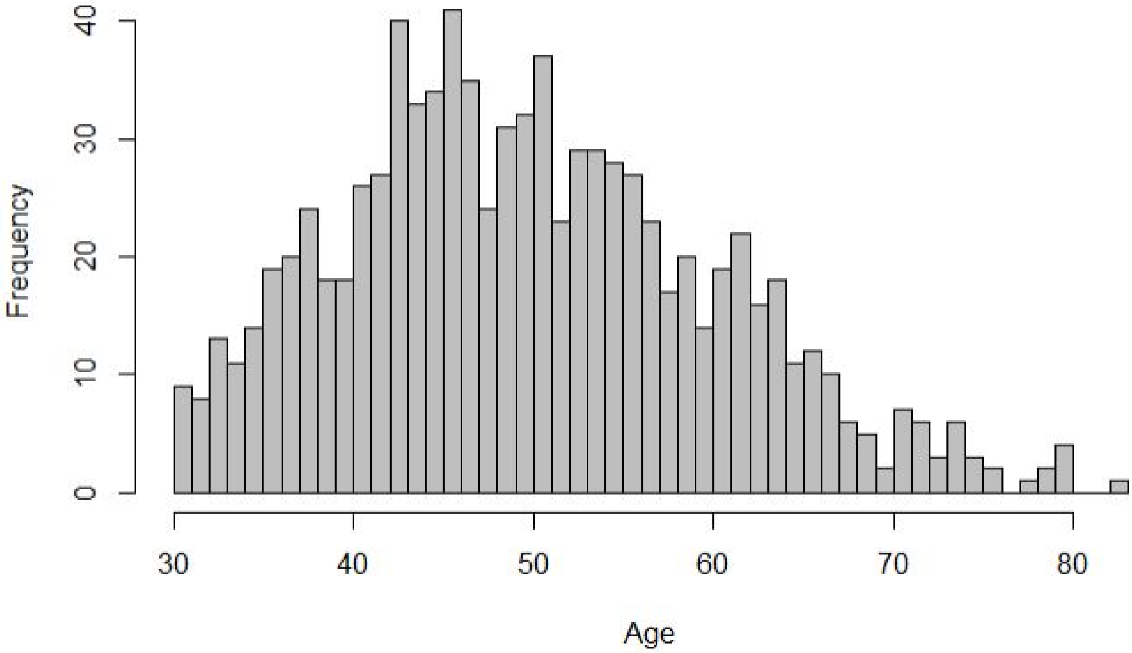


Figure 3: Histogram showing the age distribution of scholars: mean 50.2296, median 49, variance 105.8062 (standard deviation 10.2862), skewness 0.4139 (right-skewed distribution) , kurtosis -0.2380 (platykurtic distribution).

Age is expected to be a crucial factor in explaining individual human capital. In fact, as age increases, also the probability to publish great researches increases, augmenting, in turn, the individual quality index. An ad-hoc regression is run, confirming this expectation at 1% significance level (Table 9 - Appendix). Given the shape of the residuals, it does not appear necessary to include any square coefficient. Using the coefficients of the age regression on the individual quality, an age-adjusted individual human capital index is computed, predicting its value at 40 years old for each scholar. Utilizing this new index, the benchmark model is run again with the same specifications as before. Table 10 in the Appendix shows the results.

*Distance* ($ln_d$) and *agglomeration* ($ln_Pk$, $ln_Yk$, $ln_Q$) coefficients have the same sign as in the benchmark model and are still highly significant: the first hypothesis holds in the age-adjusted model, too. For what concern *selection* ($ln_q ln_d$) further investigations are needed: when the whole scholars' distribution is considered the sign of this effect is negative, although not significant. Hence, it has been decided to divide the scholars' distribution of human capital at the 95th percentile, to disclose eventual non-linear effects. The coefficients become positive only for scholars belonging to the top 5%, while they remain negative for the bottom 95%. For this reason, positive selection can be associated only with academics with really high human capital index[31]. Anyway, significance cannot be found, denying the third hypothesis also for the very top tail of the scholars' distribution. *Sorting* effect is positive even when the total

[31] Additional trials were developed to define the exact portion of scholars which is associated to positive selection and is belonging to the top tail of the distribution, however already at 90th percentile the sign becomes negative for top scholars as well.

distribution is considered ($ln_q ln_Q$), but not significant. Hence, the second hypothesis does not hold: it is not possible to claim that scholars with higher human capital index weigh more the quality of universities with respect to those with lower individual quality.

In conclusion, when age is considered the model confirms only standard results concerning distance and agglomeration, but not more sophisticated ones, like selection and sorting. These findings indicate that age was already included in the previous human capital index, which provided for lower quality levels for younger scholars and higher bibliometric indicators for senior professors. This considered, the benchmark model remains the original one, presented in Table 5.

**Gender.** In the dataset, women consist of roughly one-third of the sample (28,77%). Thus, it would be interesting to study the presence of gender differences among the effects investigated in the benchmark model.

The same regression is estimated with the addition of an interaction term: each categorical variable interacts with a dummy underlining scholars' gender (1 if male, 0 if female). Table 11 (in the Appendix) summarises the results and it is clear that none of the interaction terms with the gender dummy ($ln_d xM$, $ln_P kxM$, $ln_Y kxM$, $ln_Q xM$, $ln_q ln_d xM$, $ln_q ln_Q xM$) is significant, thus there is no evidence of gender differences. Negative *distance* coefficients remain with almost the same magnitude as in the benchmark when only women are considered, exceeding 0.8 when men are involved. Similarly, two out of three agglomeration coefficients are reinforced if male scholars are analysed. The income effect is instead greater for women, with a reduction of 0.146 for men. For what concern selection and sorting the effect is larger when the male portion of the sample is involved. Women's coefficients are almost always in line with the benchmark model, confirming that significant gender differences can not be found.

**Multiple affiliations.** After having eliminated universities with less than 5 scholars, multiple affiliation count for 3.04% of the sample. The choices of these scholars[32] are over-weighted with respect to single movers[33], given how they are considered in the dataset (see section 4). However, there should not be an over-representation of better scholars against worse ones as in the baseline research, because repeat movers are associated with a low-quality score in this dataset, with an average human capital of -0.211 against 0.169 of single movers[34]. However, to confirm the fact that repeat movers are not influencing benchmark results, in columns (3) and (4) of Table 12 (in the Appendix) they are firstly excluded from the dataset and then, in columns (5) and (6) they are associated to only one random university among their chosen ones.

[32] To be remembered: these scholars will be called 'repeated movers', meaning that they are associated to more than one university.

[33] To be remembered: with the term 'single mover' it is meant a scholar associated to only one university.

[34] The mean of $ln_q$ is computed considering 1050 dyadic matches for repeat movers (50 observations) and 16359 for single movers (779 observations).

Both modifications of the sample do not alter any result. Signs and significance are exactly the previous, while magnitude experience some changes, in particular, $ln_d$ is increased.

Furthermore, until now career choices are assumed to be independent between each other, as required by the IIA assumption in standard logistic models, although this could not be the case on many occasions. For this reason, to shed light on the eventuality of correlated preferences, a mixed logit model is developed. This version of logistic regression allows to consider the presence of heterogeneous agents. It is similar to the standard model but more flexible: the coefficients are scholars-specific and the utility function includes an additional term permitting correlated choices and the relaxation of the IIA assumption (Ye et al., 2020; Train, 2009). Technical details are exposed in the Appendix, following the results of the mixed logit estimation (columns (7) and (8) in Table 12) are illustrated.

With respect to the benchmark model, the magnitude varies for every coefficient. Signs and significance levels of *distance* coefficients are confirming gravity literature, and its magnitude increases by more than 0.3, exceeding the unity when fixed effects are considered. However, all the other results lose their significance and experience a significant decrease in magnitudes. The signs of *agglomeration* are all the opposite compared to the benchmark, but only income is slightly significant. Moreover, magnitudes of *selection* and *sorting* drastically halved. Even with the additional six parameters estimated with the mixed logit (not reported in Table 12), the LR test between columns (7) and (1) is rejecting the benchmark version. Nevertheless, the mixed logit is weaker than the benchmark since it involves simulations and not a maximization. Moreover, the assumption about parameters' distribution is essential to obtained these results, which may change when another assumption is considered.

**Private/Public universities.** As mentioned in the description of the sample (Section 4), two of the universities originally considered are private: Bocconi and Catholic University. Being privately founded implies having more autonomy on hiring processes and pay policies (Trivellato et al., 2016) to create the best context to attract relevant human capital. To understand how private institutions are influencing the benchmark estimation, another regression is run excluding both of them from the sample. The total of dyadic matches is now 11989, with 631 observations and 19 universities. The results are presented in Table 6.

*Distance* coefficients are confirming previous findings, each of them is negative, highly significant, and with a magnitude a little larger than the benchmark, but still in line with the literature of contemporary times. *Agglomeration* variables (columns (5) and (6)) represented by population size ($ln_P k$) and university notability ($ln_Q$) maintain their signs and significance levels. The magnitude of notability is greater than the one of $ln_P k$, confirming agglomeration. However, the income coefficient ($ln_Y k$) is not significant anymore, although its sign is still positive. Its magnitude decreased significantly, likely because both of the private universities excluded are located in Milan, the city with the highest disposable income. When *selection* ($ln_q ln_d$) is included in the model, the sign is positive, the significance level is at 1% and a magnitude is similar to the benchmark model. *Sorting* coefficients are positive and highly significant as well, their magnitude almost doubled. Table 6 is confirming *positive selection* and *positive sorting*, bringing evidence for a reinforcement of the latter effect. Hence, it is possible to claim that all the hypotheses hold also when only public universities are included in the model. Notwithstanding, to investigate deeper the differences between the two sectors, further developments of the project should involve a nested logit model. The current analysis can not give precise conclusions only by excluding private universities from the sample, given that error terms are still considered to be independent, as required by the standard multinomial logit model. The nested logit would allow an effective comparison of the two systems, admitting the relaxation of the IIA assumption. This method would still deny correlation between the two sectors (private and public), but now there is the possibility of error terms dependency within a nest (McFadden, 1978).

In addition to the previous regression, further insights about the influence of Bocconi and Catholic University are given in the Appendix.

Table 6: Multinomial logit regressions: standard logit model, birthplaces analysis - private universities excluded

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance:** | | | | | | |
| $ln_d$ | $-0.781^{***}$ | $-0.795^{***}$ | $-0.782^{***}$ | $-0.793^{***}$ | $-0.770^{***}$ | $-0.782^{***}$ |
| | (0.028) | (0.029) | (0.028) | (0.029) | (0.027) | (0.027) |
| **Selection:** | | | | | | |
| $ln_q ln_d$ | | $0.043^{***}$ | | $0.041^{***}$ | | $0.041^{***}$ |
| | | (0.013) | | (0.013) | | (0.013) |
| **Sorting:** | | | | | | |
| $ln_q ln_Q$ | | | $0.062^{***}$ | $0.058^{***}$ | | $0.065^{***}$ |
| | | | (0.020) | (0.019) | | (0.020) |
| **Agglomeration:** | | | | | | |
| $ln_P k$ | | | | | $-0.169^{***}$ | $-0.176^{***}$ |
| | | | | | (0.026) | (0.027) |
| $ln_Y k$ | | | | | 0.327 | 0.415 |
| | | | | | (0.430) | (0.432) |
| $ln_Q$ | | | | | $0.361^{***}$ | $0.352^{***}$ |
| | | | | | (0.046) | (0.047) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs. | 654 | 654 | 654 | 654 | 654 | 654 |
| $R^2$ | 0.249 | 0.252 | 0.252 | 0.255 | | |
| LL | −1,343.424 | −1,337.437 | −1,338.302 | −1,332.980 | −1,419.595 | −1,408.159 |
| *Note:* | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | | |

## 7 Theoretical results: present Vs past

In mapping the contemporaneous academic market in Italy several patterns have been disclosed. It is interesting to compare these characteristics with those found when the same logistic regression is run using a sample of professors who lived in Italy from 1734 to 1800[35], which is the last period of the baseline research. Agglomeration variables are not available, so the first hypothesis cannot be compared: in Table 7 the column (1) summarises the other findings using present professors, while column (2) involves past scholars.

In both cases, standard results for *distance* of gravity models are confirmed: the greater is the distance, the lower is the probability a scholar will choose to travel that route. From Table 7, the difference in magnitude between these coefficients is evident, but both are still in line with the respective literature. *Selection* effects have more similar magnitudes. Nevertheless, the coefficient for current times has a slightly greater significance, probably influenced by changes in the individual quality measures. In fact, the human quality indexes are both the result of a *PCA*, but they are considering different bibliometric indicators[36], which is affecting the results. When *sorting* is involved, another important difference between the two models must be highlighted. Notability of the university in the second column is computed aggregating the 5 highest human capital indexes associated with scholars active in that institution during the preceding 25 years (for more technical details see de la Croix and Stelter, 2021). In the first column, instead, the notability index is the one linked by RePEc to each university 10 years ago (see section 5.2). Further developments of the project could implement a version of notability index similar to de la Croix and Stelter, 2021, to avoid the disadvantages of RePEc indicators (i.e. institutions with many registered affiliated scholars are advantaged, Zimmermann, 2012) and to exclude other possible endogeneity problems. For what concerns the significance of sorting coefficients in Table 7, it is possible to claim only a weak effect in current times against a zero relevance of the same effect in the past. These results might derive from the recent reforms of the Italian academic world, as explained previously. The centralization of the system fostered equality but left aside the development of excellence (as happened all over Europe, MacLeod and Urquiola, 2021). Now, with the local recruitment of professors and the greater autonomy of each university, quality should gain attention and importance. However, these reforms are too recent to be strongly detected by the current analysis, given that there is still the influence of the previous seniority-based apparatus. All considered, this is why only a weak sorting effect is found when a sample of contemporaneous Italian scholars is analysed. The same structural

[35] This sample is provided to this project by the supervisor and responsible of the baseline research: David de la Croix.

[36] Present indicators: RePEc score, Worldcat works and library holdings, Google Scholar citations, H-index and i10-index, WoS H-index, Scopus H-index.
Past indicators: number of characters of the longest Wikipedia page, number of Wikipedia pages in different languages, Worldcat works, library holdings, and publication languages.

reason can shed light on the zero significance level of sorting during past times (column (2)): the strong control of the powers in charge had limited the relevance of universities' quality while favouring more denominational sorting, based on membership and networks rather than on meritocracy (MacLeod and Urquiola, 2021).

Table 7: Multinomial logit regressions: standard logit model, birthplaces analysis - Comparison of results from the present and from the past

| | PRESENT | PAST |
|---|---|---|
| | (1) | (2) |
| **Distance:** | | |
| $ln_d$ | $-0.794^{***}$ | $-1.707^{***}$ |
| | (0.027) | (0.058) |
| **Selection:** | | |
| $ln_q ln_d$ | $0.047^{***}$ | $0.039^{**}$ |
| | (0.012) | (0.020) |
| **Sorting:** | | |
| $ln_q ln_Q$ | $0.031^{*}$ | 0.006 |
| | (0.018) | (0.010) |
| $k$ FE | YES | YES |
| Obs | 829 | 896 |
| $R^2$ | 0.215 | 0.516 |
| LL | −1,833.845 | −1,088.157 |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

## 8 Conclusions

From a unique sample of contemporaneous scholars, this thesis confirmed and disclosed important features of the Italian academic market. Distance is highlighting a recurrent effect widely explained in migration literature, and agglomeration forces of Italian universities are present as well. The high significance of university notability is of particular interest since its positive sign shows that the quality of institutions is a strong factor for attracting academics. Selection effect is also remarkably strong in the benchmark model, implying that contemporaneous professors travel longer distances when they are associated with greater human capital indexes. Sorting is weaker in that specification, but still significant and positive, meaning that notability is more relevant for scholars with higher individual quality index when they are choosing where to develop their career. This last effect, although less clear, might be the beginning of a process initialised by recent reforms of the academic world, which may improve the quality of Italian universities by continuing to flatten inequalities, while favouring excellence.

Future research can investigate further these characteristics. Firstly, the assumption of the demand perfectly elastic in academia might be improved. This would imply the application of different models, like gravity ones, which are not posing the same strict restriction as in the multinomial logit. Moreover, gravity models allow considering both sides of the market, thus leading to a more complex, general equilibrium analysis. Secondly, remaining within logistic regressions and partial equilibrium analysis, the IIA assumption can be relaxed utilizing other versions of the same model. In this project only the mixed logit is investigated, to allow the presence of heterogeneity among the population. However, also the nested logit could be utilized. This would capture differences among groups of universities with similar characteristics. The distinction might be by status (private and public), by notability (low, middle, and high quality), or also by location (North, Centre, and South of Italy). Finally, when the standard multinomial logit model used in this thesis is considered, the notability measure can be improved. This development could be done by computing an index more similar to the one of the baseline research (de la Croix et al., 2020) in order to reduce the cons of RePEc indicators and to lower (if not to exclude) endogeneity issues.

# 9 Appendix

Table 8: Multinomial logit regressions: standard logit model - lowest level of education analysis

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance:** | | | | | | |
| $ln_d$ | $-0.735^{***}$ | $-0.735^{***}$ | $-0.735^{***}$ | $-0.735^{***}$ | $-0.728^{***}$ | $-0.728^{***}$ |
| | (0.021) | (0.021) | (0.021) | (0.021) | (0.020) | (0.020) |
| **Selection:** | | | | | | |
| $ln_q ln_d$ | | 0.012 | | 0.013 | | 0.012 |
| | | (0.009) | | (0.009) | | (0.009) |
| **Sorting:** | | | | | | |
| $ln_q ln_Q$ | | | 0.026 | 0.027 | | 0.030 |
| | | | (0.018) | (0.018) | | (0.019) |
| **Agglomeration:** | | | | | | |
| $ln_P k$ | | | | | $-0.159^{***}$ | $-0.161^{***}$ |
| | | | | | (0.026) | (0.026) |
| $ln_Y k$ | | | | | $0.977^{***}$ | $1.019^{***}$ |
| | | | | | (0.336) | (0.337) |
| $ln_Q$ | | | | | $0.219^{***}$ | $0.221^{***}$ |
| | | | | | (0.042) | (0.042) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs | 891 | 891 | 891 | 891 | 891 | 891 |
| $R^2$ | 0.278 | 0.278 | 0.278 | 0.279 | | |
| LL | −1,807.377 | −1,806.562 | −1,806.345 | −1,805.416 | −1,918.291 | −1,916.311 |
| *Note:* | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | | |

Table 9: Ordinary Least Square: age regression on human capital

| | $ln_q$ |
|---|---|
| AGE | $0.048^{***}$ |
| | (0.002) |
| Constant | $-2.281^{***}$ |
| | (0.084) |
| Observations | 16,821 |
| $R^2$ | 0.050 |
| Adjusted $R^2$ | 0.050 |
| Residual Std. Error | 2.088 (df = 16819) |
| F Statistic | $889.574^{***}$ (df = 1; 16819) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 10: Multinomial logit regressions: standard logit model, birthplaces age-adjusted analysis

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| **Distance:** | | | | | | | | | |
| $ln_d$ | −0.776*** | −0.776*** | −0.776*** | −0.777*** | −0.777*** | −0.777*** | −0.771*** | −0.771*** | −0.771*** |
| | (0.027) | (0.027) | (0.027) | (0.027) | (0.027) | (0.027) | (0.026) | (0.026) | (0.026) |
| **Selection:** | | | | | | | | | |
| $ln_q ln_d$ | | −0.004 | | | −0.002 | | | −0.004 | |
| | | (0.016) | | | (0.016) | | | (0.016) | |
| **Non-linear selection:** | | | | | | | | | |
| $top_d q$ | | | 0.010 | | | 0.013 | | | 0.019 |
| | | | (0.036) | | | (0.037) | | | (0.036) |
| $bottom_d q$ | | | −0.007 | | | −0.006 | | | −0.010 |
| | | | (0.018) | | | (0.018) | | | (0.018) |
| **Sorting:** | | | | | | | | | |
| $ln_q ln_Q$ | | | | 0.012 | 0.012 | 0.012 | | 0.008 | 0.008 |
| | | | | (0.019) | (0.019) | (0.019) | | (0.019) | (0.019) |
| **Agglomeration:** | | | | | | | | | |
| $ln_P k$ | | | | | | | −0.168*** | −0.169*** | −0.168*** |
| | | | | | | | (0.027) | (0.027) | (0.027) |
| $ln_Y k$ | | | | | | | 2.472*** | 2.474*** | 2.472*** |
| | | | | | | | (0.325) | (0.325) | (0.325) |
| $ln_Q$ | | | | | | | 0.232*** | 0.231*** | 0.231*** |
| | | | | | | | (0.042) | (0.042) | (0.042) |
| $k$ FE | YES | YES | YES | YES | YES | YES | NO | NO | NO |
| Obs | 801 | 801 | 801 | 801 | 801 | 801 | 801 | 801 | 801 |
| $R^2$ | 0.208 | 0.208 | 0.208 | 0.208 | 0.208 | 0.208 | | | |
| LL | −1,786.403 | −1,786.375 | −1,786.275 | −1,786.188 | −1,786.181 | −1,786.078 | −1,891.762 | −1,891.621 | −1,891.354 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

$top_d q$ = selection associated to the top 5% of scholars' distribution
$bottom_d q$ = selection associated to the bottom 95% of scholars' distribution

Table 11: Multinomial logit regressions: standard logit model, birthplaces analysis - gender differences

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance:** | | | | | | |
| $ln_d$ | $-0.780^{***}$ | $-0.796^{***}$ | $-0.781^{***}$ | $-0.795^{***}$ | $-0.774^{***}$ | $-0.788^{***}$ |
| | (0.027) | (0.027) | (0.027) | (0.027) | (0.026) | (0.027) |
| $ln_d xM$ | −0.028 | −0.031 | −0.029 | −0.033 | −0.039 | −0.040 |
| | (0.039) | (0.040) | (0.039) | (0.040) | (0.040) | (0.040) |
| **Selection:** | | | | | | |
| $ln_q ln_d$ | | $0.048^{***}$ | | $0.047^{***}$ | | $0.047^{***}$ |
| | | (0.012) | | (0.012) | | (0.012) |
| $ln_q ln_d xM$ | | 0.004 | | 0.006 | | 0.004 |
| | | (0.017) | | (0.017) | | (0.017) |
| **Sorting:** | | | | | | |
| $ln_q ln_Q$ | | | $0.036^{**}$ | $0.032^{*}$ | | $0.035^{*}$ |
| | | | (0.018) | (0.018) | | (0.019) |
| $ln_q ln_Q xM$ | | | 0.011 | 0.013 | | 0.008 |
| | | | (0.021) | (0.021) | | (0.021) |
| **Agglomeration:** | | | | | | |
| $ln_P k$ | | | | | $-0.173^{***}$ | $-0.178^{***}$ |
| | | | | | (0.026) | (0.026) |
| $ln_P kxM$ | | | | | −0.013 | −0.014 |
| | | | | | (0.030) | (0.030) |
| $ln_Y k$ | | | | | $2.471^{***}$ | $2.567^{***}$ |
| | | | | | (0.322) | (0.323) |
| $ln_Y kxM$ | | | | | −0.166 | −0.146 |
| | | | | | (0.334) | (0.335) |
| $ln_Q$ | | | | | $0.232^{***}$ | $0.228^{***}$ |
| | | | | | (0.042) | (0.042) |
| $ln_Q xM$ | | | | | 0.009 | 0.012 |
| | | | | | (0.052) | (0.053) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs. | 829 | 829 | 829 | 829 | 829 | 829 |
| $R^2$ | 0.210 | 0.214 | 0.211 | 0.215 | | |
| LL | −1,843.792 | −1,834.987 | −1,841.714 | −1,833.323 | −1,952.108 | −1,941.707 |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

"$xM$" represents the relative effect when only $M$ale scholars are considered.

Table 12: Repeat Movers' robustness checks and Mixed logit - birthplace analysis

| | Benchmark | | Removing RM | | RM linked to 1uni. | | Mixed Logit | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **Distance:** | | | | | | | | |
| $ln_d$ | −0.794*** | −0.787*** | −0.799*** | −0.793*** | −0.800*** | −0.793*** | −1.125*** | −0.931*** |
| | (0.027) | (0.027) | (0.028) | (0.027) | (0.028) | (0.027) | (0.074) | (0.051) |
| **Agglomeration:** | | | | | | | | |
| $ln_P k$ | | −0.178*** | | −0.177*** | | −0.184*** | | 0.013 |
| | | (0.026) | | (0.027) | | (0.027) | | (0.035) |
| $ln_Y k$ | | 2.567*** | | 2.407*** | | 2.472*** | | −0.616* |
| | | (0.323) | | (0.335) | | (0.330) | | (0.324) |
| $ln_Q$ | | 0.228*** | | 0.217*** | | 0.233*** | | −0.004 |
| | | (0.042) | | (0.044) | | (0.043) | | (0.052) |
| **Selection:** | | | | | | | | |
| $ln_q ln_d$ | 0.047*** | 0.047*** | 0.051*** | 0.050*** | 0.051*** | 0.050*** | 0.024 | 0.021 |
| | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) | (0.020) | (0.019) |
| **Sorting:** | | | | | | | | |
| $ln_q ln_Q$ | 0.031* | 0.035* | 0.037* | 0.041** | 0.033* | 0.037** | 0.014 | 0.016 |
| | (0.018) | (0.019) | (0.019) | (0.019) | (0.018) | (0.019) | (0.021) | (0.021) |
| $k$ FE | YES | NO | YES | NO | YES | NO | YES | NO |
| Obs | 829 | 829 | 779 | 779 | 804 | 804 | 829 | 829 |
| $R^2$ | 0.215 | | | 0.231 | | | | |
| LL | −1,833.845 | −1,942.488 | −1,693.084 | −1,813.279 | −1,748.797 | −1,874.075 | −1,795.738 | −1,988.388 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

RM = Repeat Movers

Note: The mixed logit involves the six s.d. associated to each coefficient, only the s.d. linked to $ln_d$ is significantly different from zero.

**Mixed Logit Model.** The mixed logit model is applied to count for the heterogeneity of individuals and to permit the relaxation of the IIA assumption. In particular, coefficients are not fixed anymore but they are varying across the population and an additional term in the utility function allows for correlated choices (Train, 2009; Ye et al., 2020). Hence, in the mixed logit model the scholar's utility function from moving towards a university $k$ is modified as follows:

$$U_{ik} = \beta_i x_{ik} + \eta_i x_{ik} + \epsilon_{ik} \tag{6}$$

Where the first term is the general deterministic component, representing the utility of scholar $i$ choosing university $k$. The other two terms are capturing the unobservable part of the function, $\eta_i$ is an individual deviation, and $\epsilon_{ik}$ is a random term as before. These two error terms are assumed to be normally distributed.

In the mixed logit model, with the $\eta$ term violating the IIA assumption, it is required the integration of the conditional probability using the joint probability density function, $f(\beta_i|\theta)$; where $\theta$ summarise the first and the second moment of the distribution. It is assumed that the vector of $\beta$ coefficients is independent and normally distributed and it is of length $N$. To obtain the unconditional probability that professor $i$ is choosing university $k$, the following formula is applied:

$$P_{ik} = E(P_{ik}|\beta_i) = \int_{\beta} (P_{ik}|\beta_i) f(\beta_i|\theta) d\beta = \int_{\beta_1} \int_{\beta_2} \dots \int_{\beta_N} (P_{ik}|\beta_i) f(\beta_i|\theta) d\beta_1 d\beta_2 \dots d\beta_N \tag{7}$$

where

$$(P_{ik}|\beta_i) = \frac{exp(\beta_i x_{ik})}{\sum_{k' \in K} exp(\beta_i x_{ik'})} \tag{8}$$

is the conditional probability.

Simulations are used to draw the parameters from the $\beta$ distribution: the unconditional probability is the average of the conditional probabilities, which are computed for each scholar (Train, 2009; Ye et al., 2020; Leon, 2013). Table 13 is presenting all the specifications of the mixed logit regression, columns (4) and (6) are those reported in Table 12.

**Private/Public universities - further insights.** In addition to the regression presented in the text, other two are developed in this paragraph: one excluding only Catholic University while keeping Bocconi, and another excluding Bocconi while keeping Catholic University. Thanks to the former, it will be possible to investigate the issue of secondary locations, understanding whether it alters previous results. The latter, instead, will examine the position of excellence reached by Bocconi in several rankings. In fact, it has lately become the best uni-

Table 13: Multinomial logit regressions: mixed logit model - birthplaces analysis

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance** | | | | | | |
| $ln_d$ | $-1.124^{***}$ | $-1.124^{***}$ | $-1.124^{***}$ | $-1.125^{***}$ | $-0.930^{***}$ | $-0.931^{***}$ |
| | (0.074) | (0.074) | (0.074) | (0.074) | (0.051) | (0.051) |
| **Selection** | | | | | | |
| $ln_q ln_d$ | | 0.022 | | 0.024 | | 0.021 |
| | | (0.019) | | (0.020) | | (0.019) |
| **Sorting** | | | | | | |
| $ln_q ln_Q$ | | | 0.008 | 0.014 | | 0.016 |
| | | | (0.021) | (0.021) | | (0.021) |
| **Agglomeration** | | | | | | |
| $ln_P k$ | | | | | 0.016 | 0.013 |
| | | | | | (0.035) | (0.035) |
| $ln_Y k$ | | | | | $-0.672^{**}$ | $-0.616^{*}$ |
| | | | | | (0.320) | (0.324) |
| $ln_Q$ | | | | | 0.003 | $-0.004$ |
| | | | | | (0.050) | (0.052) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs | 829 | 829 | 829 | 829 | 829 | 829 |
| $R^2$ | 0.230 | 0.231 | 0.231 | 0.231 | | |
| LL | −1,797.090 | −1,796.013 | −1,796.337 | −1,795.738 | −1,989.772 | −1,988.388 |
| *Note:* | | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | | |

versity in Italy for economics and related fields and this may be due to its well-known ability to attract high-ranked personalities.

The results obtained by excluding Catholic University can be seen in Table 14: almost all coefficients are significant. *Distance* and *agglomeration* remain as in the benchmark (Table 5) for what concern signs and significance, the magnitude is pretty similar as well, with income coefficient experiencing the largest variation: a drop of 0.271 (from 2.567 to 2.296). *Positive selection* maintains its significance level of 1% in each specification. *Positive sorting* is again weaker than selection, but it gains some significance in the third and sixth column with respect to the benchmark. This implies that there is no relevant bias due to the imprecise geographical coordinates associated with this university. Although every scholar is assumed to teach only in Milan, the benchmark model is not significantly influenced. However, further developments of the project may improve this aspect.

Excluding Bocconi from the set of choices allows finding only significant coefficients, as demonstrated in Table 15, improving the solidity of the results. *Distance* and all *agglomeration* variables remain with the same sign and significance as in the benchmark model. The magnitude of the coefficient associated with university notability increases, while the income one is almost halved. *Positive selection* is found also here, as in the previous estimation without Catholic University. Furthermore, in this regression, there is strong evidence for *positive sorting* in each specification. The high significance level reached with this model means that private universities have different features, which are not totally captured by the variables included in the benchmark: further investigations are needed (i.e. applying different models). Despite this, it is possible to claim that all the expected features of the contemporaneous academic world described in Section 3 are still holding if Bocconi University is excluded.

Table 14: Multinomial logit regressions: standard logit model, birthplaces analysis - Catholic University excluded

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance:** | | | | | | |
| $ln_d$ | $-0.774^{***}$ | $-0.788^{***}$ | $-0.775^{***}$ | $-0.787^{***}$ | $-0.769^{***}$ | $-0.781^{***}$ |
| | (0.027) | (0.028) | (0.027) | (0.028) | (0.026) | (0.027) |
| **Selection:** | | | | | | |
| $ln_q ln_d$ | | $0.046^{***}$ | | $0.044^{***}$ | | $0.043^{***}$ |
| | | (0.012) | | (0.012) | | (0.012) |
| **Sorting:** | | | | | | |
| $ln_q ln_Q$ | | | $0.040^{**}$ | $0.035^{*}$ | | $0.039^{**}$ |
| | | | (0.019) | (0.019) | | (0.019) |
| **Agglomeration:** | | | | | | |
| $ln_P k$ | | | | | $-0.167^{***}$ | $-0.172^{***}$ |
| | | | | | (0.026) | (0.027) |
| $ln_Y k$ | | | | | $2.194^{***}$ | $2.296^{***}$ |
| | | | | | (0.347) | (0.348) |
| $ln_Q$ | | | | | $0.238^{***}$ | $0.232^{***}$ |
| | | | | | (0.042) | (0.042) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs | 755 | 755 | 755 | 755 | 755 | 755 |
| $R^2$ | 0.225 | 0.229 | 0.226 | 0.230 | | |
| LL | −1,615.601 | −1,608.114 | −1,613.318 | −1,606.342 | −1,722.265 | −1,712.776 |
| *Note:* | | | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ | | |

Table 15: Multinomial logit regressions: standard logit model, birthplaces analysis - Bocconi University excluded

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Distance:** | | | | | | |
| $ln_d$ | $-0.779^{***}$ | $-0.795^{***}$ | $-0.781^{***}$ | $-0.793^{***}$ | $-0.772^{***}$ | $-0.786^{***}$ |
| | (0.027) | (0.028) | (0.027) | (0.028) | (0.026) | (0.027) |
| **Selection:** | | | | | | |
| $ln_q ln_d$ | | $0.046^{***}$ | | $0.044^{***}$ | | $0.043^{***}$ |
| | | (0.012) | | (0.012) | | (0.012) |
| **Sorting:** | | | | | | |
| $ln_q ln_Q$ | | | $0.059^{***}$ | $0.054^{***}$ | | $0.060^{***}$ |
| | | | (0.019) | (0.019) | | (0.019) |
| **Agglomeration:** | | | | | | |
| $ln_P k$ | | | | | $-0.175^{***}$ | $-0.182^{***}$ |
| | | | | | (0.026) | (0.026) |
| $ln_Y k$ | | | | | $1.350^{***}$ | $1.448^{***}$ |
| | | | | | (0.366) | (0.368) |
| $ln_Q$ | | | | | $0.318^{***}$ | $0.311^{***}$ |
| | | | | | (0.044) | (0.044) |
| $k$ FE | YES | YES | YES | YES | NO | NO |
| Obs | 728 | 728 | 728 | 728 | 728 | 728 |
| $R^2$ | 0.232 | 0.235 | 0.234 | 0.237 | | |
| LL | −1,558.063 | −1,550.982 | −1,553.242 | −1,546.893 | −1,643.476 | −1,631.674 |
| *Note:* | | | | $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ | | |

# Bibliography


Agasisti, T., Barra, C., & Zotti, R. (2019). *Research, knowledge transfer, and innovation: The effect of Italian universities' efficiency on local economic development 2006/2012.* (Vol. 59). Journal of Regional Science.

Agasisti, T., & Bianco, A. D. (2007). *Determinants of college student migration in italy: Empirical evidence from a gravity approach.*

Akcigit, U., Baslandze, S., & Stantcheva, S. (2016). *Taxation and the international mobility of inventors.* (Vol. 106/10). American Economic Review.

Audretsch, D. (1998). *Agglomeration and the location of innovative activity.* (Vol. 14/2). Oxford Review of Economic Policy.

Baldissera, A., & Cornali, F. (2020). *Geography of human capital in italy: A comparison between macro-regions.* (Vol. 25). Cambridge University Press.

Barone, C., & Guetto, R. (2016). *Verso una meritocrazia dell'istruzione? Inerzia e mutamento nei legami tra origini sociali, opportunità di studio e destini lavorativi in Italia (1920–2009).* (Vol. 1).

Barra, C., & Zotti, R. (2017). *Investigating the human capital development–growth nexus does the efficiency of universities matter?* (Vol. 40/06). International Regional Science Review.

Barra, C., & Zotti, R. (2018). *The contribution of university, private and public sector resources to italian regional innovation system (in)efficiency.* The Journal of Technology Transfer.

Barrientos, P. (2007). *Analysis of international migration and its impacts on developing countries.* (Vol. 12). Institute for Advanced Development Studies, Development Research Working Paper.

Barro, R. (2001). *Human capital and growth.* (Vol. 91/2). American Economic Review, American Economic Association.

Beine, M., Docquier, F., & Özden, Ç. (2011). *Diasporas.* (Vol. 95).

Bertola, G., & Sestino, P. (2011). *A comparative perspective on italy's human capital accumulation.* (Vol. 6). Economic History Working Papers, Bank of Italy.

Bini, M., & Chiandotto, B. (2003). *La valutazione del sistema universitario italiano alla luce della riforma dei cicli e degli ordinamenti didattici.* (Vol. 2). Studi e note di economia.

Bratti, M., & Verzillo, S. (2019). *The 'gravity' of quality: Research quality and the attractiveness of universities in italy.* [Regional Studies].

Capano, G. (2008). *Looking for serendipity: The problematical reform of government within italy's universities.* (Vol. 55).

Checchi, D., Fraja, G. D., & Verzillo, S. (2014). *Publish or perish: An analysis of the academic job market in italy.* (Vol. 14/04) [Discussion Paper]. University of Nottingham, School of Economics.

Checchi, D., & Verzillo, S. (2014). *Selecting university professors in italy: Much ado about nothing?* Mimeo, Milano Italy.

Cottini, E., Ghinetti, P., & Moriconi, S. (2019). *Higher education supply, neighbourhood effects and economic welfare.* [DISCE - Working Papers del Dipartimento di Economia e Finanza]. Università Cattolica del Sacro Cuore, Dipartimenti e Istituti di Scienze Economiche (DISCE).

Croissant, Y. (2020). *Multinomial logit models.*

de la Croix, D., Docquier, F., Fabre, A., & R., S. (2020). *The academic market and the rise of universities in medieval and early modern europe.*

de la Croix, D., & Stelter, R. (2021). *Scholars and Literati at the University of Göttingen (1734–1800).* Repertorium Eruditorum Totius Europae - RETE.

Docquier, F., & Marfouk, A. (2006). *International migration by educational attainment (1990–2000).* [In: Ozden, C., Schiff, M. (Eds.), International Migration, Remittances and Development.]. Palgrave Macmillan, New York. Chapter 5.

Drucker, J., & Goldstein, H. (2007). *Assessing the regional economic development impacts of universities: A review of current approaches.* (Vol. 30). International Regional Science Review.

Durante, R., Labartino, G., & Perotti, R. (2011). *Academic dynasties: Decentralization and familism in the italian academia.* [NBER Working Paper No. 17572].

Faggian, A., & McCann, P. (2009). *Universities, agglomerations and graduate human capital mobility.* (Vol. 100/2). Journal of Economic; Social Geography.

Grogger, J., & Hanson, G. (2016). *Income maximization and the selection and sorting of international migrants.* (Vol. 95/1). Journal of Development Economics.

Grogger, J., & Hanson, H. (2015). *Attracting talent: Location choices of foreign-born phds in the united states.* (Vol. 35(S1)). Journal of Labor Economics, University of Chicago Press.

Handler, H. (2018). *Economic links between education and migration: An overview.* (Vol. 4) [Flash Paper]. Policy Crossover Center Vienna - Europe.

Jacso, P. (2005). *As we may search–comparison of major features of the web of science, scopus, and google scholar citation-based and citation-enhanced databases.* (Vol. 89). Current Science.

Kerr, S., Kerr, W., Özden, Ç., & Parsons, C. (2016). *Global talent flows.* (Vol. 30/4). Journal of Economic Perspectives, American Economic Association.

Kerr, S., Kerr, W., Özden, Ç., & Parsons, C. (2017). *High-skilled migration and agglomeration.* (Vol. 9). Annual Review of Economics.

Leon, G. (2013). *Transportation choices and the value of statistical life.* (Vol. 9).

MacLeod, W. B., & Urquiola, M. (2021). *Why does the united states have the best research universities? incentives, resources, and virtuous circles.* (Vol. 35/1).

McFadden, D. (1974). *Condition logit analysis of quantitative choice behavior.* University of California at Berkeley. Chapter 4.

McFadden, D. (1978). *Modeling the choice of residential location.* Transportation Research Record.

Mokyr, J., Sarid, A., & VanDerBeek, K. (2020). *The wheels of change: Human capital, millwrights, and industrialisation in 18th-century england.* Voxeu.

Morana, M. (2020). *Focus "il personale docente e non docente nel sistema universitario italiano - a.a 2019/2020.* Elaborazioni su banche dati MIUR, DGCASIS – Ufficio VI Gestione patrimonio informativo e statistica.

Nelson, R., & Phelps, E. (1966). *Investment in humans, technological diffusion, and economic growth.* (Vol. 56(1/2)). The American Economic Review.

Neuhaus, C., & Hans-Dieter, D. (2008). *Data sources for performing citation analysis: An overview.* (Vol. 64).

Norris, M., & Oppenheim, C. (2007). *Comparing alternatives to the web of science for coverage of the social sciences' literature.* Infometrics.

Ortega, F., & Peri, G. (2013). *The effect of income and immigration policies on international migration.* (Vol. 1). Migration Studies.

Perotti, R. (2008). *L'università truccata.* Einaudi. Torino, Italy.

Ramos, R. (2016). *Gravity models: A tool for migration analysis.* Institute of Labor Economics (IZA).

Rebora, G., & Turri, M. (2008). *La governance del sistema universitario in italia: 1989-2008.* (Vol. 32) [The World of Economics]. Liuc Papers Serie Economia Aziendale.

Schiller, D., & Cordes, A. (2016). *Measuring Researcher Mobility – A Comparison of Different Datasets and Methods with an Empirical Application of Micro-Data for the United States and Germany.*

Schwartz, A. (1976). *Migration, age and education.* (Vol. 24). Journal of Political Economy.

Seiler, C., & Wohlrabe, K. (2011). *Ranking economists and economic institutions using repec: Some remarks.*

Squicciarini, M., & Voigtländer, N. (2015). *Human capital and industrialization: Evidence from the age of enlightenment.* (Vol. 30). Quarterly Journal of Economics.

Stephan, P., & Levin, S. (2001). *Exceptional contributions to us science by the foreign-born and foreign-educated.* (Vol. 20(1/2)). Population Research; Policy Review.

Train, K. E. (2009). *Discrete choice methods with simulation.* (2nd ed.). Cambridge University Press.

Trivellato, P., Triventi, M., & Traini, C. (2016). *Private higher education in italy.*

Triventi, M., & Trivellato, P. (2008). *Le onde lunghe dell'università italiana. partecipazione e risultati accademici degli studenti nel novecento.* (Vol. 22). Polis (Italy).

Williams, R. (2019). *Panel data 3: Conditional logit/ fixed effects logit models.* University of Notre Dame.

Yang, K., & Meho, L. (2006). *Citation analysis: A comparison of google scholar, scopus, and web of science.* (Vol. 43). Proc. Am. Soc. Info. Sci. Tech.

Ye, M., Chen, Y., Yang, G., Wang, B., & Hu, Q. (2020). *Mixed logit models for travelers' mode shifting considering bike-sharing.* (Vol. 12).

Zhao, X., Aref, S., Zagheni, E., & Stecklov, G. (2021). *International migration in academia and citation performance: An analysis of german-affiliated researchers by gender and discipline using scopus publications 1996-2020.*

Zimmermann, C. (2012). *Academic rankings with repec.* [Working Papers 2012-023]. Federal Reserve Bank of St. Louis.