



Universidad Autónoma
de Madrid

Biblos-e Archivo
Repositorio Institucional UAM

Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:
This is an **author produced version** of a paper published in:

Historical Methods 55.1 (2022): 45-60

DOI: <https://doi.org/10.1080/01615440.2021.2014376>

Copyright: © 2021 Taylor & Francis Group, LLC.

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

"This is an Accepted Manuscript version of the following article, accepted for publication in Historical Methods 55.1 (2022): 45-60. It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited."

EconHist: a relational database for analyzing the evolution of economic history (1980–2019)

Abstract: Since the cliometric revolution, the future of economic history has been discussed in relation to its supposedly increasing integration with economics and other disciplines. Any well-grounded argument in this regard would require a quantitative and qualitative analysis of the scientific production of economic historians in recent decades. This article provides a systematic method for collecting and analyzing the scientific production—in the form of indexed articles—of a broad and representative sample of authors who identify themselves as economic historians. From this sample, we have built EconHist, a relational database that contains the bibliometric information provided by Scopus, and the biographical information from authors' curricula vitae between 1980 and 2019. Finally, we show the opportunities and difficulties related to the design and development of such a database.

Keywords: relational databases, bibliometric analysis, economic history, economic historians

I. Introduction

The inherent interdisciplinarity of economic history makes it a cross-discipline between large domains of knowledge or even a field subject to a clash of disciplines (Jerven, 2011). Cipolla provided one of the most accurate definitions of the discipline when describing the field as “an eminently interdisciplinary subject: it occupies an area of human knowledge at the crossroads of two other disciplines, history and economics, and cannot afford to ignore either of them. If it yields ground on either front, it becomes distorted and runs the risk of losing its identity” (Cipolla, 1992, pp. ix–x).

In the last years, the alleged integration of economic history into economics has received increasing attention (Cioni, Federico, & Vasta, 2018, 2020, 2021; Hauptert, 2019; Jaremski, 2020; Margo 2021). Although there is not a single definition or metric for assessing that integration, the idea is that economic history is increasingly converging to the methods, type of publications, and language usually employed in economics.

Some authors have celebrated economic history’s leaning towards economics. Romer, for example, noted that “the field of economic history is no longer a separate, and perhaps marginal, subfield of economics, but rather, is an integral part of the entire discipline” (Romer, 1994, p. 49). She provides several examples of how the discipline has enriched our understanding of the economic past and economics in general. Diebolt and Hauptert also note that “[b]y merging economic history with modern techniques, cliometricians have not ended economic history, but elevated it” (2018a, p. 431).

Nevertheless, Margo, a leading figure in the cliometric movement that approached economic history to economics since the 1960s, recognized that the “consequences of the integration of economic history into economics may not be wholly positive” (Margo, 2018, p. 398). Other authors have expressed similar concerns. For example, in the passage above, Cipolla warned against the dangers of ignoring any of the disciplines at the core

of economic history. Many authors saw the increasing use of quantitative methods in economic history brought by the cliometric revolution as a “colonization” of the discipline (Godden, 2013). Under this more pessimistic view, the discipline would be consigned to a mere subfield of applied economics and lost both its transversality and the rigor in data-gathering and the critical analysis of historical sources or the data quality. Another fierce critique of the cliometric revolution, Boldrizzoni, also sees the integration of economic history into economics as an attempt “to create narratives o the past compatible with neoliberal economics, (...) and to endorse specific worldviews, theories, and policy recommendations.” Boldrizzoni considers that economic history requires “a new pact between history and the social sciences in order to counter the way economists have abused the past.” In his view, “we need a different paradigm of historical research that is not subject to economic theory but contributes towards renewing it” (Boldizzoni, 2011: 5-6).

The relationship of economic history with economics and history might also impact scholars’ training and academic curricula. While economic history is being published in top economic journals and complements the education of economists (Jaremski, 2019), the reliance on quantitative methods and techniques typical of economists might reinforce the cultural turn in history. In the extreme, this could result in an (almost) complete dissociation between scholars trained in history and those working in economic history. The recent controversies between historians of capitalism and economic historians around slavery illustrate the dangers of this tendency. After revisiting the debates surrounding *Time on the Cross* since its publication, Hilt concluded that “most of the new works by historians on slavery have not engaged with the literature that emerged in the wake of *Time on the Cross*, and in some cases, those books are vulnerable

to some of the same critiques” (Hilt, 2020, p. 479). The dissociation between economic history and history also has sensible academic costs.

Rather than focusing on the optimistic or pessimistic views above or measuring the academic costs of economic history distancing itself from history, this article takes a step back to better measure the degree to which the integration of economic history in economics is (or is not) a reality. We build EconHist, a database to study the demographic, academic, and scientific trends in economic history. In our view, claims about the costs (or promises) of the integration of economic history into economics need ample evidence on the practitioners’ profiles and work to test their validity or consistency. EconHist provides a unique and rich source of information for the profile of economic historians and their scientific output during the last forty years.

Some authors working on the relationship of economic history with other disciplines (especially economics) have used bibliometric and networking studies for leading journals in the field (Galofré-Vilà, 2020), articles published (Morgan & Shanahan, 2010; Wehrheim, 2018; Whaples, 1991, 2002), topic classification (Cherrier, 2017), or topic analysis using structural topic models (stm).¹ These works typically involve building tables or databases based on the information downloaded from the Web of Science (WoS), Scopus, EconLit, or Google Scholar. This way of proceeding is not always mechanical and may even become, as we will see, very time-consuming.

Despite their merits and advantages, these investigations are, in our view, partial, and to some extent, biased or incomplete. Most of these works start from a very restrictive selection of the sample of authors or works they study, like a journal (or 3-5 top journals) or a group of highly cited researchers. Herein, we adopt a different strategy. We start from a more general characterization of economic historians, which identifies a broader and

¹ For stm methodology, see Roberts, Stewart, and Tingley (2016). An application in Economic History is Grajzl and Murrell (2020). See also Blei (2012).

more representative sample of practitioners—at least for Europe and North America—and analyze their scientific output and the relationship with other biographical characteristics.²

Our bibliometric analysis adopts a revealed-preference approach to identify the work produced by economic historians since 1980. An economic historian is defined as someone who explicitly manifests an interest in economic history by being a member of at least one leading association or attending a major conference in economic history and has published at least one peer-reviewed article in an economic history journal indexed in Scopus.³ This characterization is, of course, subjective and partial because it leaves out many economic historians who do not appear in the lists of associations or have not published in an economic history journal. The associations and conferences that we use also lead to an overrepresentation of European and North American scholars and underrepresents scholars from other continents. However, it allows us to obtain a broader and representative sample of practitioners and their scientific output and collaborators. In this sense, our selection criterion surmounts—at least partially—the selection bias pointed out above and provides a much broader population and a much more varied scientific production for a better understanding of the path taken by the profession.

We have built a relational database for this sample of economic historians that links the scientific production with the authors' biographical characteristics. In other words, we have combined the authors' scientific production in Scopus⁴ with their biographical information obtained from publicly available online sources such as curricula vitae (CV), personal or official websites, and other open online sources. This

² According to Baten and Muschallik (2012, p. 110) by 2012 there were around 10,400 scholars in the world. Our sample includes biographical information of 1108 authors (around 10.5% of the population).

³ Economic History Association, the Cliometric Society, the Economic History Society, and the European Historical Economics Society. We exclude Ph.D. students from the sample.

⁴ The reasons why we use this bibliographic database are explained in the next section.

database is generated by crossing information about articles, authors, affiliations, and sources with the data of 1108 authors. Of them, 876 (79%) authors had published at least one peer-reviewed article indexed in Scopus. The database contains 70 variables and 10,773 bibliometric records and the references cited in those articles (155,960). The database provides an excellent basis for analyzing and studying co-citation networks, co-authors, major source magazines, publications since 1980, production by region (distinguishing mainly between North America and Europe), or trends by gender or cohort.

This study introduces the advantages and difficulties of building referential databases for researching the evolution of an academic field—economic history.⁵ While our research is directly related to the growing literature on the impact of the bibliometric revolution and the integration of economic history into economics, our approach provides at least two original contributions: first, it overcomes some of the limitations of previous approaches. Second, it provides broader and more varied information that allows for richer analyses than those existing to date, thereby shedding more light on recent trends in economic history.

The article is organized as follows. Section 2 describes how we have built the EconHist database, its main features, and its limitations. Section 3 compares our database with the ones used in similar articles published in the last decade. This section also describes relevant outcomes and explores some trends in the discipline using the information in the database. Finally, we present the problems, limitations, and opportunities arising from our methodology and offer some practical recommendations to facilitate and improve this type of research.

⁵ On the importance of such databases for historical-economic research, see Perez-Garcia (2019).

2. Materials and methods

2.1 Data collection

Data collection followed two main steps. First, we defined the sample of economic historians. Second, we collected both bibliometric and biographic data for that sample.

2.1.1 The population

Our list of practitioners was based on the memberships to the Economic History Association (EHA) and the Cliometric Society (CS), and participants at the 2018 and 2019 meetings of the Economic History Society (EHS) and the 2017 meeting of the European Historical Economics Society (EHES). EHA provided the names for 994 practitioners, whereas CS provided the names for 289 practitioners. Finally, participants at the EHS and EHES meetings provided 237 and 144 names of new practitioners, respectively.⁶ In total, our initial sample contained 1664 practitioners.⁷ After eliminating individuals for whom no individual information was found on the internet or had not been awarded a Ph.D. degree by mid-2019, the final sample contains 1108 individuals. Of those, 876 (79%) have at least one article indexed in Scopus.⁸

When using membership or participation in conferences organized by Anglo-Saxon or European organizations as the primary criterion for being included in our sample, the database has a marked Western bias. Almost 90% of the individuals in the

⁶ 579 individuals (34.8%) appeared in more than one the four lists that we used for our sample.

⁷ We are indebted to Mike Hauptert for providing the list of members of the EHA and the CS. Participants at the 2018 EHS meeting can be found at <https://ehs.org.uk/wp-content/uploads/2020/11/Conference-Booklet-2018.pdf>. Participants at the 2019 meeting can be found at <https://ehs.org.uk/wp-content/uploads/2020/11/Conference-Booklet-2019.pdf>. Participants at the 2017 EHES meeting are available at https://uni-tuebingen.de/uploads/media/EHES_2017_Programme_01.pdf. All links were accessed on December 1, 2020.

⁸ See section 2.1.3 below for more details.

sample are affiliated with European or North American organizations, and 96% earned their doctorate from a university in these continents.⁹ Although geographically biased, the organizations that we consider are among the leading ones in economic history, as shown by the fact that they edit some of the top journals in the discipline.¹⁰ Nevertheless, the reader must remember that our subsequent results mainly apply to economic history as practiced in Europe and North America.

2.1.2 The bibliometric production

Once the authors had been selected, we searched their scientific production in the form of journal articles indexed in Scopus. There are two reasons for restricting our search to peer-reviewed articles in Scopus. The first reason is that, contrary to the articles published in journals, other publications (such as books or book chapters) tend to be perceived as a different channel for consolidated information instead of cutting-edge research in economics and other disciplines.¹¹ The second reason is technical: both WoS and Scopus provide metadata that allows for a straightforward bibliometric analysis by using, for example, network analysis software, such as VOSviewer, Pajek, BibExcel, HisCite, Sci², and CiteSpace. However, although the data from WoS and Scopus sometimes overlap (Mongeon and Paul-Hus, 2016), WoS presents some significant limitations given its regional scope. WoS has an important bias toward Anglo-Saxon and German countries (Di Vaio, Waldenström and Weisdorf, 2012). On the contrary, Scopus has more extensive

⁹ See section 3.2. below for a more detailed analysis of descriptive statistics. Most of individuals working in North America (430 out of 478 or 89.96%) do it in the United States.

¹⁰ In this sense, our work is comparable for example to that by Di Vaio and Weisdorf (2009), which includes the three journals published by EHA, EHS and EHEH in the top 4 major journals in economic history.

¹¹ There is an additional reason. Although for academic book authors and the institutions assessing their research performance, the relevance of books is undisputed, the absence of comprehensive international databases covering the items and information needed for the assessment of this type of publication imposes a severe limitation. Several European countries are developing custom-built information systems for the registration of scholarly books, as well as weighting and funding allocation procedures (see Giménez-Toledo et al. (2016)).

coverage of European and Asian journals, which facilitates comparing the scientific production of American and European economic historians. We are aware that other databases such as Google Scholar or ResearchGate also include many references. To some extent, they are even more comprehensive than WoS or Scopus.¹² However, these databases have problems (such as the reliability of the information or its structure), which makes Scopus a better option.¹³

2.1.3 The CV database

The biographical information for the individuals in our database was obtained using publicly available online information. When the CV was not available, we used alternative sources of information such as professional websites, articles, blog posts, or other pieces discussing the individual's career. After omitting the data of those individuals from the initial 1664 for whom no information was available online or had not been awarded a Ph.D. degree by August 2019, we were left with data of 1,108 individuals (66.59%) with at least some partial demographic and academic information. We will be using the data of 1,108 individuals using the database's demographic and academic characteristics. Of them, 876 (79.06%) had at least one publication in a peer-reviewed journal related to economic history indexed in Scopus. This subgroup constituted our sample for further analyses of their scientific production.

Table 1 shows the distribution of individuals by decade when they were awarded their doctoral degrees.¹⁴

¹² For a comparison between Google Scholar, WoS, and Scopus based on citations in 252 subject categories, see Martín-Martín et al. (2018).

¹³ There are works that already use these data sources, but their use requires prior programming and an enormous information debugging work. See, for example, Martín-Martín, Costas, van Leeuwen, and López-Cózar (2018).

¹⁴ The number of observations is less than 1114 because not all the individuals have information for the year when they were awarded their Ph.D. degree.

[Table 1 about here]

Individuals awarded their Ph.D. degrees in the last two decades constituted around 60 percent of our sample. There are at least two reasons for this. First, the members of older cohorts were more likely to have left academia either due to pursuing other jobs or because they retired or passed away. Second, the number of doctoral degrees awarded annually has been increasing over time.¹⁵ Our database does offer a representative sample of all the economic historians since the cliometric revolution in the late 1950s–early 1960s. It presents a highly representative and detailed sample of *current* practitioners in economic history and their scientific production in the form of articles.

2.2 *The database*

2.2.1 Main features

From data collection, we built a referential database with 70 variables and 10,773 records. The variables have been grouped into four tables: the first table shows characteristics of the authors, the second shows the articles they have published from 1980 to January 2020,¹⁶ the third shows their current affiliations and, the fourth table concerns journals. Appendix I shows the descriptions of the tables and variables. Herein, we highlight some of the descriptions that will be useful for some of the analyses in later sections. We present them according to the tables in the database.

¹⁵ In the US, for example, the number of doctoral recipients increased by almost 65 percent between 1988 and 2018 (see Table I2 from the Survey of Earned Doctorates by the National Science Foundation available at <https://nces.nsf.gov/pubs/nsf20301/data-tables/>; accessed on November 12, 2020).

¹⁶ Previously to 1980, information in Scopus (and in WoS) is scattered and rather incomplete.

Table *Article*

We distinguish between the order of author names in the publication (author_1 through author_5) because the order of authorship may be relevant in some cases. We excluded articles with six or more authors (0.67% of total articles).¹⁷ Table 2 shows the distribution of articles by the number of authors. More than three-fourth (75.7%) of the articles have one or two authors.¹⁸ Articles with four or more authors only represent 7.67% of the total articles.

[Table 2 about here]

The articles in our database received 249,064 citations, with an average of 25.88 citations per document (standard deviation 74.99).¹⁹ The variable *references* includes 155,960 works—mainly journal articles, but also books, book sections, working papers, and reports.

Table *Author*

This table shows the primary contributions of this study. Bibliometric data is well-structured and can be processed straightforwardly with databases such as Scopus and

¹⁷ Two papers had more than 30 authors!

¹⁸ Seltzer and Hamermesh (2018) report that the average number of authors per paper in the main journals range from 1 to 2.19. Co-authorship could be used as a multidisciplinary proxy. In principle, such a high rate of single authorship restrains multidisciplinary. Usually, when multiple authors collaborate, the probability that they come from different disciplines is higher. Moreover, collaborations seek to take advantage not only from synergies but also from co-authors' different skills. It is common to see, for example, an economic historian publishing together with an econometric specialist and vice versa. In the field of history, there is even fewer co-authorships than in economic history.

¹⁹ The highest number of citations (2850) corresponds to Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. (1980). Geometry from a time series. *Physical Review Letters*, 45(9): 712–716. This article is neither about economic history nor is published in an Economic history journal. However, according to his CV, one co-author (J.D. Farmer) has been identified as economic historian because he participated in one EHS conference. The second most cited paper (2067) is North, D.C. and Weingast, B.R. (1989). Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth Century England. *Journal of Economic History*, 49(4): 803–832.

WoS. Still, the process comes with some challenges when referring to authors' data. Quite often, biographical data lacks a structure and is spread among several sources and formats. Using the scholars' CV, we extracted binary variables for the membership to EHA or CS or attendance to EHS or EHES meetings, gender,²⁰ place of work (country and type of department), professional rank, year of Ph.D. graduation, university and country that granted the doctoral degree, and Ph.D. discipline. When available, we also included their fields of interest and the year of birth.

Tables *Journal* and *Affiliation*

The first table simply adds information about the sources (*journal_name*) and other publishing information. The second table refers to the organization, country, and region where researchers have a position (usually held during our online search in the second and third quarters of 2019 or before retirement). These data facilitate regional comparisons, especially between Europe and North America (the two regions better represented in our database).

[Figure 1 about here]

²⁰ The majority of CV or other biographical documents do not contain explicit information about gender. We used Gender API algorithm (<https://gender-api.com>; accessed November 11, 2020) to infer gender. This algorithm gives a probabilistic estimate for the person's gender based on their first name. One important limitation of this approach is that it imposes a binary structure for gender (male or female); therefore, our imputed gender does not necessarily reflect the actual gender identity. Furthermore, the association of a given name with a given gender may vary across countries. This will typically result in a low first-name-based gender probability (e.g. the algorithm gives "Andrea" a 54%-chance of being female probably because it is commonly used for females in English-speaking countries, but in Italy, it is commonly used for males). We searched additional online information for individuals with a first-name-based gender probability below 60%. When public online documents for those individuals used pronouns that did not correspond to the first-name-based gender prediction, the imputation was based on those pronouns. For example, Andrea Papadia's profile at the European University Institute's website states: "*He* completed *his* PhD at the London School of Economics" (emphasis added); therefore, the gender for this person was changed to "Male" (see <https://www.eui.eu/ProgrammesAndFellowships/MaxWeberProgramme/People/MaxWeberFellows/Fellows-2017-2018/Papadia>; accessed November 11, 2020).

The database has been developed in SQL and can be managed in several ways using MS Access, Excel, or Power Pivot.²¹ Figure 1 shows the structure of the EconHist database. In addition, as a step prior to elaborating the bibliographic table (which includes *article* and, partially, *Author*), Scopus generates a .csv file that can conveniently be used with most network analysis software. In our case, we used VOSviewer and Pajek.²²

2.2.2 Some problems

Although powerful databases like Scopus debug, maintain, and structure much of the information, many difficulties persisted and had to be addressed in an almost artisanal and labor-intensive way. For example, filters used by Scopus lead to some errors. Despite filtering by document type to extract only those published before January 2020, some book reviews (and even book sections and reports) were obtained in the search result. Such errors are probably caused due to misclassifications in the encoding of the document. Duplication is another issue: despite the low number of records used compared to large databases, we identified more than two hundred duplicated items.²³

The preparation of the data also involved the standardization of journal names. For example, *The American Economic Review* may appear as “The American Economic Review,” “American Economic Review,” “Amer. Econ. Rev.,” “Amer Econ Rev,” or simply “AER.” We used full names by eliminating the article “the” when it appears at the beginning (so, in the previous example, the journal name always becomes “American Economic Review”).²⁴ In cases where the journal name appears in English and other

²¹ Power Pivot is an app that runs over MS Excel.

²² VOSviewer is a free software for visualizing scientific landscapes (such as co-citation networks). For further details, see van Eck and Waltman (2014). Pajek is a free software for network analysis (see De Nooy, Mrvar, and Batagelj, 2005).

²³ On some occasions, we found the same article in two different issues of the same journal (with different pagination) or translations of the same work into different languages.

²⁴ Thus, “Journal of Economic History,” “Economic Journal,” etc., instead of “The Journal of Economic History,” “The Economic Journal,” etc.

languages, we used the English name unless the journal is officially identified in the other language.²⁵ Standardizing the name of the sources (not only for economic history journals) was also very important to maintain a consistent identification of the sources cited in the articles recorded in our database.²⁶

We encountered similar issues with authors' names. In addition to the unsystematic use of the initials, there were inconsistent spellings involving tildes, commonly used in Spanish and French; the "ö" and "ü" in German; and the numerous Slavic, Scandinavian, or Turkish characters. We used UTF-8 encoding and normalized, whenever possible, spellings of names by removing tildes and other special characters. In addition, people with compound surnames (common in Spanish) were identified multiple times as different authors depending on how the surname has been listed.²⁷ We systematized several hundred authors' names and surnames by removing tildes and other special characters.

Indexing author names deserves a separate discussion. Using a single identification code (e.g., ORCID or ResearcherID) for each author would solve or significantly mitigate the previous problems with author names. However, sometimes, ID codes render the identification of authors more complicated if one author has several identification codes. For example, Leandro Prados de la Escosura has been recorded in three different ways: "Prados de la Escosura, Leandro" (Scopus ID: 55982545400, 9 records), "De la Escosura, Leandro Prados" (Scopus ID: 6505827385, 26 records), "Escosura, Leandro Prados de la" (Scopus ID: 36652659400, only 2 records), and "Prados

²⁵ For example, in the case of the "Revista de Historia Económica, Journal of Iberian and Latin American Economic History," we maintained only "Revista de Historia Economica."

²⁶ See Online Appendix A for other checks and corrections that we made related to journals, their names, and misclassifications in Scopus.

²⁷ For a database it is not the same "Fernández de Pinedo," which is the right form in Spanish, and not "Fernandez-de-Pinedo," which is incorrect in Spanish. However, the use of hyphen was extended as a way of preventing the surname from being truncated in an international journal—which may make a difference in the case of a surname as common in Spanish as Fernández.

de la Escosura, Leandro” (Scopus ID: 8141705600, 2 records). Cormac Ó Gráda presents similar issues. There are many more cases in which the author appears with two codes — corresponding to two different ways to be indexed. We had to identify and unify in a single register all the different namings for a given author.

Other issues arose for unknown reasons. For example, there was a varying degree of completeness in the records—many of them had blank fields or fields for which there is no information. We had to regularize the positions of fields across all records to make the data arrays uniform.

The part of the database related to the authors’ biographical information also presented some significant challenges. The heterogeneity of formats in the sources (personal or professional websites, or LinkedIn), the lack of electronic CVs, or the poorly standardized or incomplete CVs forced us to manually build much of the 33 fields for the 1,108 biographical documents that we collected.

All of the process required for retrieving and cleaning the data involves a considerable amount of back-office time. In addition, if this process is delegated to a research assistant, clear guidance and constant monitoring on the complex and numerous decisions for every step—for example, ensuring that the articles are attributed to the correct researcher—is recommended.²⁸

3. Results

²⁸ For example, it has not always been easy to disambiguate the production of authors with very common names for whom we did not have a CV or had an incomplete one. An example of this issue happened with authors with common surnames of Asian origin such as Lee. The author’s search yielded hundreds of results in Scopus and it was difficult to identify the precise author that we were looking for.

This section presents some examples of different approaches and empirical and analytical tools to exploit the information in the database.

3.1 A comparison to other databases

Table 3 summarizes other databases on some topics that can be addressed with our database (e.g., the integration of economic history into economics). Many of these other works do not rely on databases as such.

[Table 3 about here]

Most of the studies have used tables or databases that focus mainly—or exclusively²⁹—on *The Journal of Economic History*, or the so-called “Top 5” journals. EconHist is more comprehensive than the studies that focus on top journals because it includes more information (10,773 records)³⁰ and does not select only some journals but takes all the authors’ peer-reviewed publications indexed in Scopus. Another advantage of EconHist with respect to the works listed in Table 3 is that it provides detailed authors’ biographical information. Only Abramitzky (2015) and Baten and Muschallik (2012) included some (partial) information on biographical data. Finally, EconHist is also a highly reliable database because issues such as repetitions or ambiguities concerning authors’ names and other types of publication errors have been revised, and inconsistencies have been eliminated. This consistency is more difficult to achieve in other works and databases that rely on massive data collection.

²⁹ For example, Whaples (1991, 2002).

³⁰ Only surpassed by Angrist et al. (2017) which, however, is a massive and much more generalist database—conditions needed to apply machine learning techniques.

3.2 Descriptive analysis

Individual's biographical information provides several promising venues to explore the integration of economic history into economics and the changes that have taken place within economic history over time. For example, at least 56 percent of the individuals in the database obtained their Ph.D. in Economics versus less than 14 percent in History (Table 4).³¹

[Table 4 about here]

We can also analyze authors' distribution by gender and authors' "productivity" in terms of journal articles (Figures 2 and 3).

[Figure 2 about here]

Figure 2 suggests that women are under-represented in economic history. Even if the proportion of female scholars has been increasing in the most recent cohorts, they represent less than one-fourth (around 23.5 percent) of the individuals in our sample. Women are less than 20 percent among those who obtained their Ph.D. before 1980 and about 25 percent for those who graduated after 2000. Our numbers are remarkably similar to what Hale and Regev found for graduates in the top 10 US Economics departments between 1987 and 2007 (Hale and Regev, 2014, p. 66). Our results are also consistent

³¹ The 42-point gap is larger when accounting for the fact that the "Other" category in Table 4 includes individuals with doctoral degree in the field related to economic (e.g., "Economics and Business," "Development Economics," or "Agricultural Economics"). There are also some instances of doctoral degrees in the field related to history (e.g. "History and Politics" or "History and Theology"). After creating two large categories for Ph.D. programs in "Economics and Finance" and "History," the former includes more than 62 percent of individuals in our database, whereas the later includes 16 percent (a 46-point gap). See Online Appendix B for the criteria that we followed to group Ph.D. programs into broader categories for economics, history, and economic history.

(although slightly lower) with the ones presented by Bayer and Rouse when using a more comprehensive sample of US Economics doctorates (Bayer & Rouse, 2016). These authors find that around 30 percent of the doctorate degrees were awarded to women between 1995 and 2014.³² The gender gap, which is receiving increasing attention in Economics, is also a reality among economic historians.³³

[Figure 3 about here]

Figure 3 shows an interesting pattern for the average articles per capita production per decade after graduation. The horizontal axis represents the decades since the Ph.D. degree was awarded (for example, “Decade 0” for someone graduating in 1954 would be the 1950s).³⁴ The vertical axis shows the average number of articles per capita published by members of each cohort of doctorates in that decade from their graduation. As expected, the curves show an inverted-U shape: articles per capita increase during the first 2-3 decades after graduation and start declining around the fifth decade. Interestingly, the most recent generations of doctors (especially the ones being awarded a doctorate in the last 20 years) have concentrated much of their scientific production in the form of indexed articles, and they seem to be on average more productive as decades go by. For example, the 2000s cohort published more than eight articles per capita in their first decade after graduation (the 2010s), significantly higher than the 5-6 articles per capita published during the 1970s, 1980s, and 1990s cohorts in their first decade since graduation.

³² All the mentioned works use people being awarded a Ph.D. from economics departments, whereas our database is mostly made of individuals who pursued (and stayed) an academic career. Research suggests that women are significantly less likely to obtain tenure than men even after accounting for differences in the year of graduation or institutional quality of alma mater (Ginther and Kahn, 2014, 2004). If this is true, we should expect to find a larger gender gap in our database.

³³ For studies and initiatives to address the gender gap in economics, see for example the work of the Committee on the Status of Women in the Economics Profession at the American Economic Association at <https://www.aeaweb.org/about-aea/committees/cswep> (accessed on November 11, 2020).

³⁴ Our sample in Figure 3 is restricted to those individuals with at least one publication indexed in Scopus.

In addition to an increase in the (article) productivity for younger cohorts, the results could also be explained by changes in the incentives for promotion that might reward peer-reviewed articles more than other publications excluded from our database (e.g., books).

The rise in co-authorship in Economic History also contributes to the increase in articles per capita across cohorts. Seltzer and Hamermesh (2018) find an increase in the number of authors in the top journals in Economic History and attribute the trend to the increasing integration of economic history into economics. The results in our database confirm Seltzer and Hamermesh's findings. The average number of authors per article starts at 1.46 in the 1980s and steadily increases from 1.63 in the 1990s to 2.14 in the 2010s.

3.3 Network analysis

An attractive option in our database—especially concerning the debate about the future of economic history—is the development of network analysis. Many applications have specifically been designed to perform this type of analysis (Pajek or CitNet). Other tools are easily programmable or have plug-ins or tools for running network analysis (MatLab, R, or Stata). We have used VOSviewer, a valuable and easy-to-use tool for quickly visualizing networks (represented by nodes and their links) and clusters (using colors). Figures 4 and 5 show, respectively, the sources (journals) and the authors cited by individuals in EconHist.

[Figure 4 about here]

These two figures reinforce the idea of integration (or dependence?) of economic history into economics. Figure 4 shows that most of the works cited by the authors

appeared in the most important economic history journals (green cluster): *The Journal of Economic History* and *Economic History Review*. However, some journals shown in the figure are more revealing, like the red cluster around *The American Economic Review* and *The Economic Journal*. The other clusters are mainly associated with publications from money, finance, and banking. In sum, Figure 4 shows the enormous weight, centrality, and density of major mainstream American economic journals in economic historians' citation networks.

[Figure 5 about here]

Finally, Figure 5 shows the network and clusters of citations of prominent authors—those with more than 175 citations. The figure contains five clusters of authors. In the yellow cluster, there are topics covered by authors such as Eichengreen, Temin, Bordo, and Krugman, who have worked on economic crises, the gold standard, or the Great Depression. The central red cluster includes authors like Williamson or O'Rourke, whose research focuses on the effects of trade and factor supply in connection with the data sets compiled by Maddison or Mitchell since the 1980s. In addition, the blue cluster includes authors such as North, Weingast, Shleifer, or Acemoglu, who study the role of institutions in long-run growth. In the green cluster, authors such as Mokyr, Allen, van Zanden, Clark, and Humphries have worked on the causes of the Industrial Revolution. Finally, the purple cluster is linked to industrialization, focusing on anthropometry and health outcomes (including authors such as Komlos or Steckel). All these clusters and their relative position (centrality) give us an idea of the dominant topics in 1980-2019.³⁵

³⁵ Other possible analyses with these techniques are those of topics (from abstracts and keywords). Due to space limitations, we will not present them here.

3.4 An econometric model for economic historians' Ph.D. discipline

One empirical application for our database has to do with the econometric analysis to estimate changes among practitioners in economic history over time while holding constant other changes in the discipline. Take the idea that economic history is increasingly integrated into economics. Many authors have tested this hypothesis by looking at the publications of economic historians in the leading economics journals or the use of econometric language in their work (see Table 3). EconHist allows for a different test (that also uses a broader sample of practitioners than the one traditionally used in other works) by looking at the likelihood of practitioners in economic history being awarded a doctorate from an Economics department.

Consistently with economic history being more integrated into economics, about two-thirds of the individuals in our sample obtained a Ph.D. in Economics or Finance.³⁶ However, the percentage of individuals with an Economics Ph.D. does not vary much by the decade of graduation (see Figure 6). The trends seem flat or even slightly negative. This lack of positive would seemingly suggest that the cliometric revolution increased the integration of economic history into economics in its early years (the late 1950s to the early 1960s), but, at least since the 1980s, there is no discernible tendency toward greater integration.

The problem with that naïve inference from Figure 6 is that other factors that affect the likelihood of getting a Ph.D. in Economics might have changed over time. Take, for example, women's presence in the field of economics. The proportion of women graduating each decade has shown a slow (but steady) increase (see figure 2). If, as some evidence suggests, women are less likely to opt for economics as an academic career than

³⁶ See Online Appendix B for a list of the Ph.D. titles that were considered to be in Economics or Finance.

men, then the increasing proportion of women graduates by cohort would be driving down the percentage of practitioners with a Ph.D. in Economics.³⁷

[Figure 6 about here]

Other institutional factors could also affect the proportion of people graduating from the economics departments. For example, economic history has been disappearing from the curriculum in most doctoral programs in the United States (Temin, 2013). There are numerous departments of Economic History in Europe, but they do not exist in the United States. Changes in the relative importance of graduates from Europe or North America could influence economic historians' likelihood of receiving an Economics Ph.D. The relative importance of European and North American graduates in our database varied across cohorts. Figure 7 shows that almost 80% of pre-1970 graduates have a degree from a North American institution, whereas less than 20% graduated from a European university. Among those that graduated in the 2010s, the majority (53.54%) graduated in Europe, whereas only 40% attended a North American university.³⁸

[Figure 7 about here]

Self-selection into different societies and their meetings is another factor that can affect the likelihood that individuals in our sample have an economic background

³⁷ As shown in section 3.2, women typically represent 30 percent or less of people being awarded an Economics Ph.D., but they represent more than 40% of doctoral degrees awarded in History in the last 20 years (see the data from the American Academy of Arts & Sciences available at <https://www.amacad.org/humanities-indicators/higher-education/gender-distribution-degrees-history#31653>; accessed on November 11, 2020).

³⁸ Out of 522 individuals who were awarded a Ph.D. from a North American university in our database, 551 (94.74%) studied in the United States.

(degree). The Cliometric Society, for example, is “an academic organization of individuals interested in the use of economic theory and statistical techniques to study economic history.”³⁹ The economic history Society “exists to support research and teaching in economic and social history.”⁴⁰ Due to the very different nature of these two associations, the former is more likely to attract individuals with training in economics. In contrast, the latter seems more susceptible to including people with a degree in history.

A test of economic history’s integration into economics that examines whether there is a greater tendency for practitioners to get an economics degree must hold all those other factors that may also affect the individuals’ field for their Ph.D. constant. This is what an econometric test makes possible. In particular, we use the following regression:

$$Pr(Econ_PhD_i = 1) = F(\alpha + \beta_1 Year_i + \beta_2 Male_i + \theta \cdot Continent_PhD_i + \lambda \cdot Affiliations_i) + \varepsilon_i \quad (1)$$

Where $Econ_PhD_i$ is a dummy variable that equals one if individual i received a Ph.D. in Economics or Finance and 0 otherwise; $Year_i$ reflects the year of graduation for individual i ; $Male_i$ is a dummy variable that equals one if i is male and 0 otherwise; $Continent_PhD_i$ is a vector of dummies that equals 1 for the continent in which i obtained the Ph.D. and 0 for the rest; $Affiliations$ is a vector of associations and congresses in economic history that equals 1 for the ones that individual i attended or held membership and 0 otherwise.

Equation (1) formalizes what we described above: the probability of practitioner i obtaining a Ph.D. in Economics will be a function of the passage of time, but also of other

³⁹ <https://www.cliometrics.org/about/> (accessed on October 6, 2020).

⁴⁰ <https://www.ehs.org.uk/the-society/about-the-ehs.html> (accessed on October 6, 2020).

factors such as gender, the continent where the individual studied, or memberships in associations.

The coefficient of interest is the one for *Year*. A statistically significant value for β_1 would suggest that after controlling for the many compositional changes across cohorts in the database and the differences among associations, there is increasing integration of economic history into economics via an increasing probability of practitioners being awarded a Ph.D. in economics. A null or negative coefficient for β_1 would indicate that such integration does not exist or that our sample's economic historians are becoming less and less likely to have an economics background.

[Table 5 about here]

We estimate equation (1) using a Probit model.⁴¹ Table 5 shows the average marginal effects for the probit model. The likelihood of economic history practitioners obtaining a doctoral degree in economics significantly increases over time by 0.4 points per year. In other words, holding everything else constant, an individual awarded a Ph.D. in 2019 is almost 20 points more likely to graduate in economics than an individual who graduated in 1970.⁴² The integration of economic history into economics is alive and ongoing. The other coefficients also have the expected signs and effects for the discipline's institutional factors and compositional changes. Men are 6.6 points more likely than women to graduate from an economics department. Geographically,

⁴¹ Probit models are used for regressions in which the dependent variable (in this case, type of Ph.D.) can only adopt two values (e.g. "economics" or "not economics"). It assumes that the probability of a positive outcome is determined by the standard normal cumulative distribution function. See Online Appendix A3 for alternative specifications that do not change the results.

⁴² Results do not change when running separate regressions for people being granted their doctoral degree in Europe and in North America (as a matter of fact, the coefficient for *Year of PhD* is identical to the one reported in column 4 of Table 5 in both separate regressions). We thank an anonymous referee for suggesting this robustness check. Results available upon request.

individuals with a North American Ph.D. are 27 points more likely to graduate from an economics department than someone from Europe (the reference category in the regression). Finally, affiliations with organizations and attendance at congresses reflect the organization's character to which individuals adhere. Members of the CS are 23.2 points more likely to hold an Economics Ph.D., whereas those attending the meetings of the EHS are almost 13 points less likely to have an Economics Ph.D.⁴³

The previous results provide only some examples of using a well-defined relational database for the research questions of this study. Of course, there are alternatives both for methodology and the formulation of questions concerning the future of economic history.

4. Discussion

In this study, we built a relational database that combines bibliographic information with biographical data of economic historians. EconHist is unique thanks to its extension, content, and structure to the best of our knowledge. Although some of the references cited in this article (for example, Seltzer, 2018) extend databases (or tables) beyond the standard fields included in Scopus, they do not build relational databases. Thanks to its design and characteristics, EconHist facilitates more detailed research on the evolution of economic history after the cliometric revolution and the second revolution discussed by Cioni et al. (2021).

We are aware of EconHist's limitations. Despite our efforts to compile a representative sample of practitioners in economic history (at least in Europe and North

⁴³ Results are also robust to different specifications for the dependent variable (i.e. using narrower definition of a Ph.D. in Economics or Finance specified in Online Appendix B instead of the broadest one). If, for some reason, people without a Ph.D. in Economics retire earlier and at greater rates than people being awarded a Ph.D. in Economics, our estimates for β_1 would be upwardly biased. Excluding people with pre-1980 doctorates does not substantially affect our results either. Results for all robustness checks are shown in Online Appendix C.

America), the sample does not cover certain “obscure areas” in the profession. Many authors have not registered themselves in any association belonging to economic history. When they do, these associations sometimes have a marked “local” character—typically from their country or region of origin— and are not part of our database. Moreover, EconHist focuses exclusively on peer-reviewed articles and excludes other formats such as books, review articles, or book sections. Large bibliographic databases —mainly Scopus and WoS— and the most popular tools for exploiting them are more suitable to articles. The need to develop better tools for the compilation and analysis of the scientific production beyond peer-reviewed articles is a recurrent topic in international scientometrics forums, especially in the social sciences and the humanities.

We wish to conclude this article with some recommendations derived from our experience in developing the database. In particular, having access to updated and homogenous CVs would greatly facilitate the collection of biographical information. To the extent that CV formats depend on several factors—such as regulations from universities or national science systems—there is no systematic, consistent, or structured information about authors (if there is any). ORCID and Publons, two international initiatives, could help in systematizing this type of information.⁴⁴ Works like ours, studying researchers’ academic careers, would greatly benefit from the more systematic and updated information that these platforms could provide if used widely. One potential barrier to achieving this outcome is that the cost of keeping profiles up-to-date would fall

⁴⁴ ORCID (<https://orcid.org>) is a “nonprofit organization helping create a world in which all who participate in research, scholarship and innovation are uniquely identified and connected to their contributions and affiliations, across disciplines, borders, and times.” Publons (<https://publons.com>) is part of the WoS Group and is powered by integrations with the WoS, ORCID, and thousands of scholarly journals; this platform serves researchers and publishers. Other platforms such as editorialmanager.com or manuscriptcentral.com require the creation of a user account or being registered with ORCID. All links were accessed on November 19, 2020.

on the users, whereas the database (which, in cases like Publons, are commercial websites) would receive most of the benefits.⁴⁵

References

- Abramitzky, R. (2015). Economics and the Modern Economic Historian. *The Journal of Economic History*, 75(4), 1240-1251. doi:10.1017/S0022050715001667
- Angrist, J., Azoulay, P., Ellison, G., Hill, R., & Lu, S. F. (2017). Economic Research Evolves: Fields and Styles. *American Economic Review*, 107(5), 293-297. doi:10.1257/aer.p20171117
- Bisin, A., & Federico, G. (Eds.) (2021). *The Handbook of Historical Economics*. London: Academic Press.
- Baten, J., & Muschallik, J. (2012). The Global Status of Economic History. *Economic History of Developing Regions*, 27(1), 93-113. doi:10.1080/20780389.2012.682390
- Bayer, A., & Rouse, C. E. (2016). Diversity in the Economics Profession: A New Attack on an Old Problem. *Journal of Economic Perspectives*, 30(4), 221-242. doi:10.1257/jep.30.4.221
- Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM*, 55(4), 77-84. doi:10.1145/2133806.2133826
- Boldizzoni, Francesco, 2011. *The Poverty of Clio: Resurrecting Economic History*, 1st edition. ed. Princeton University Press, Princeton, NJ.
- Cherrier, B. (2017). Classifying Economics: A History of the JEL Codes. *Journal of Economic Literature*, 55(2), 545-579. doi:10.1257/jel.20151296

⁴⁵ On the pros and cons of this kind of practices, see Mansell and Steinmueller (2020), especially Chapter 2.

- Cioni, M., Federico, G., & Vasta, M. (2018). Ninety years of publications in Economic History: evidence from the top five field journals (1927–2017). *Quaderni del Dipartimento di Economia Politica e Statistica (Università di Siena)*(1240).
- Cioni, M., Federico, G., & Vasta, M. (2020). The long-term evolution of economic history: evidence from the top five field journals (1927–2017). *Cliometrica*, 14(1), 1–39. doi:10.1007/s11698-019-00186-x
- Cioni, M., Federico, G., & Vasta, M. (2021). The two Revolutions in Economic History. In A. Bisin & G. Federico (Eds.), *Handbook of Historical Economics*: Elsevier North Holland.
- De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory Network Analysis with Pajek*. New York: Cambridge University Press.
- Di Vaio, G., Waldenström, D., & Weisdorf, J. (2012). Citation success: Evidence from economic history journal publications. *Explorations in Economic History*, 49(1), 92–104. doi:https://doi.org/10.1016/j.eeh.2011.10.002
- Diebolt, C., & Hauptert, M. (2018a). A cliometric counterfactual: what if there had been neither Fogel nor North? *Cliometrica*, 12(3), 407–434. doi:10.1007/s11698-017-0167-8
- Diebolt, C., & Hauptert, M. (2018b). We are Ninjas: How Economic History has Infiltrated Economics. *Preliminary draft prepared for presentation at the American Economic Association meetings*.
- Fourie, J. (2019). Who Writes African Economic History? *Economic History of Developing Regions*, 34(2), 111–131. doi:10.1080/20780389.2019.1639500
- Galofré-Vilà, G. (2020). The Past’s Long Shadow. A Systematic Review and Network Analysis of Economic History. *Research in Economic History*, 36.

- Giménez-Toledo, E., Mañana-Rodríguez, J., Engels, T. C. E., Ingwersen, P., Pölönen, J., Sivertsen, G., . . . Zuccala, A. A. (2016). Taking scholarly books into account: current developments in five European countries. *Scientometrics*, 107(2), 685-699. doi:10.1007/s11192-016-1886-5
- Ginther, D.K., Kahn, S., 2014. Academic women's careers in the social sciences, in Lanteri, A., Vromen, J.J. (Eds.), *The Economics of Economists: Institutional Setting, Individual Incentives and Future Prospects*. Cambridge University Press, New York, pp. 285–315.
- Ginther, D.K., Kahn, S., 2004. Women in Economics: Moving Up or Falling Off the Academic Career Ladder? *J. Econ. Perspect.* 18, 193–214.
<https://doi.org/10.1257/0895330042162386>
- Godden, C., 2013. In praise of Clio: Recent reflections on the study of economic history. *Æconomia*, 3-4, 645-664.
- Grajzl, P., & Murrell, P. (2020). Characterizing a legal–intellectual culture: Bacon, Coke, and seventeenth-century England. *Cliometric*. doi:10.1007/s11698-020-00202-5
- Hale, G., Regev, T., 2014. Gender ratios at top PhD programs in economics. *Econ. Educ. Rev.* 41, 55–70. <https://doi.org/10.1016/j.econedurev.2014.03.007>
- Hamermesh, D. S. (2018). Citations in Economics: Measurement, Uses, and Impacts. *Journal of Economic Literature*, 56(1), 115-156. doi:10.1257/jel.20161326
- Harzing, A.-W. (2017). *Publish or Perish*. Retrieved from <https://harzing.com/resources/publish-or-perish>
- Hauptert, M. (2019). History of cliometrics. In C. Diebolt & M. Hauptert (Eds.), *Handbook of cliometrics*. Berlin-Heidelberg: Springer.

- Hilt, Eric, 2020. Revisiting Time on the Cross After 45 Years: The Slavery Debates and the New Economic History. *Capitalism: A Journal of History and Economics* 1, 456–483. <https://doi.org/10.1353/cap.2020.0000>
- Jaremski, M. (2020). Today's economic history and tomorrow's scholars. *Cliometrica*, 14, 169–180. doi:10.1007/s11698-019-00188-9
- Jerven, M. (2011). A Clash of Disciplines? Economists and Historians Approaching the African Past. *Economic History of Developing Regions*, 26(2), 111-124. doi:10.1080/20780389.2011.625244
- Mansell, R., & Steinmueller, W. E. (2020). *Platform Economics*. Cheltenham UK: Edward Elgar.
- Margo, R. A. (2018). The integration of economic history into economics. *Cliometrica*, 12, 377–406. doi:10.1007/s11698-018-0170-8
- Margo, R. A. (2021). The Economic History of Economic History: The Evolution of a Field in Economics. In A. Bisin & G. Federico (Eds.), *Handbook of historical economics*. Amsterdam and New York: Elsevier North-Holland.
- Martín-Martín, A., Costas, R., van Leeuwen, T., & López-Cózar, E. (2018). Evidence of Open Access of scientific publications in Google Scholar: a large-scale analysis. *Journal of Informetrics*, 12(3), 819-841.
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160-1177. doi:<https://doi.org/10.1016/j.joi.2018.09.002>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228. doi:10.1007/s11192-015-1765-5

- Morgan, S., & Shanahan, M. (2010). The supply of Economic History in Australasia: The Australian Economic History Review at 50. *Australian Economic History Review*, 50(3), 217-239. doi:10.1111/j.1467-8446.2010.00303.x
- Perez-Garcia, M. (2019). Consumption of Chinese goods in southwestern Europe: a multi-relational database and the vicarious consumption theory as alternative model to the industrious revolution (eighteenth century). *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 52(1), 15-36. doi:10.1080/01615440.2018.1523695
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). stm: R Package for Structural Topic Models. *Journal of Statistical Software*.
- Romer, C. D. (1994). The End of Economic History? *Journal of Economic Education*, 25(1), 49-66. doi:10.2307/1182896
- Seltzer, A. J. (2018). Publication Trends and Future Challenges for the Australian Economic History Review: A Bibliometric Analysis. *Australian Economic History Review*, 58(2), 112-133. doi:10.1111/aehr.12143
- Seltzer, A. J., & Hamermesh, D. S. (2018). Co-authorship in economic history and economics: Are we any different? *Explorations in Economic History*, 69, 102-109. doi:https://doi.org/10.1016/j.eeh.2018.04.001
- Temin, P. (2013) The Rise and Fall of Economic History at MIT. In *MIT Department of Economics Working Paper No. 13-11, Available at SSRN: https://ssrn.com/abstract=2274908 or http://dx.doi.org/10.2139/ssrn.2274908.*
- van Eck, N. J., & Waltman, L. (2014). Visualizing Bibliometric Networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact: Methods and Practice* (pp. 285-320). Cham: Springer International Publishing.

- Wehrheim, L. (2018). Economic history goes digital: topic modeling the Journal of Economic History. *Cliometrica*. doi:10.1007/s11698-018-0171-7
- Whaples, R. (1991). A Quantitative History of the Journal of Economic History and the Cliometric Revolution. *The Journal of Economic History*, 51(2), 289-301.
doi:10.1017/S0022050700038948
- Whaples, R. (2002). The Supply and Demand of Economic History: Recent Trends in the Journal of Economic History. *The Journal of Economic History*, 62(2), 524-532. Retrieved from <http://www.jstor.org/stable/2698190>

Appendix I: Database tables and variables

Table Article (25 variables)

Auth_1, Auth_2, Auth_3, Auth_4, Auth_5, Title, Year, Source_title, Volume; Issue, Art._No., Page_start, Page_end, Cited_by, DOI, Link, Abstract, author_keyword_1, author_keyword_2, author_keyword_3, author_keyword_4, author_keyword_5, Index Keywords, References

Table Author (33, 32 variables)⁴⁶

author_name, eha, clio, ehese, ehs, Margo, researcherid, orcid, gender, organization, department, rank, yearbirth, yearPh.D., universityPh.D., countryPh.D., Ph.D.program, babsyear, babsprogram, researchinterest_1, researchinterest_2, researchinterest_3, researchinterest_4, researchinterest_5, updated, noteissue, decade_Ph.D., core1, core2, core3, id_scopus_1, id_scopus_2, id_scopus_3

Table affiliation (5, 3 variables)

affiliation, *organization*, *department*, country, region_work

Table Source (12, 10 variables)

journal_name, classification_1, classification_2, publisher, ISSN, journal_URL, country, SJR, JCR, journal_field_1, journal_field_2, journal_field_3, journal_field_4

(12, 10)

⁴⁶ In this case (as well as in the next two tables), the first figure refers to the number of variables that the table contains, and the second one how many *new* variables that table adds to the database. In italics characters the variables also included in other tables.

Figures



Figure 1. The structure of EconHist relational database.

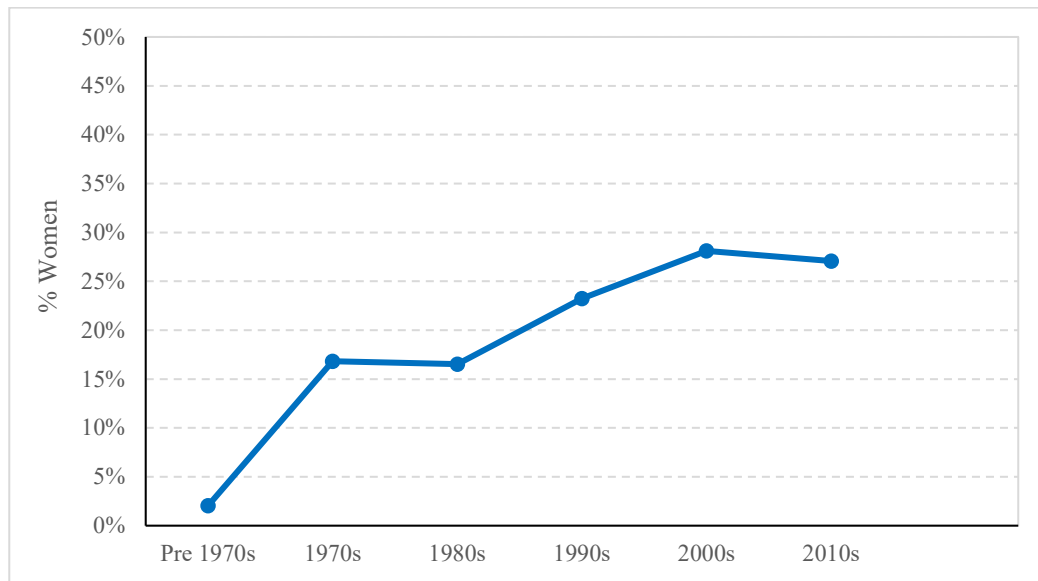


Figure 2. Share of doctoral degrees awarded to women by decade in EconHist database.

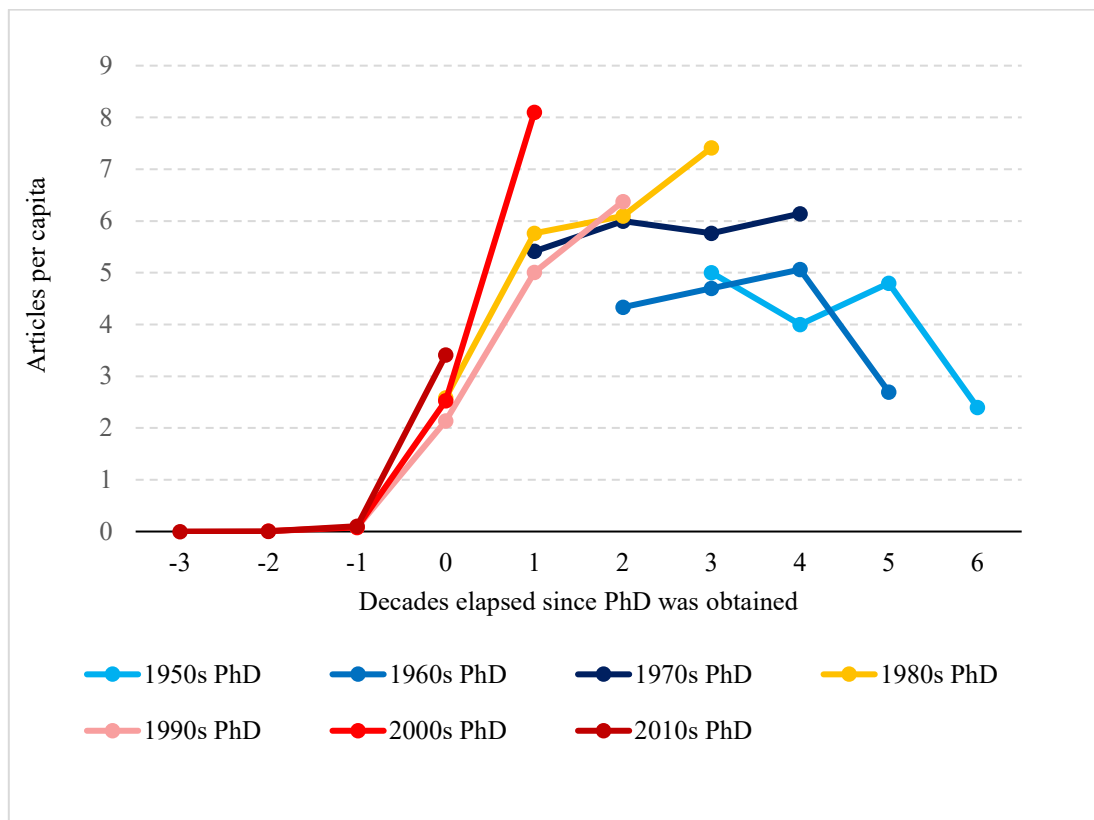


Figure 3. Evolution of the number of articles published by Ph.D. decade for scholars with at least one publication in an Economic History journal indexed in Scopus.

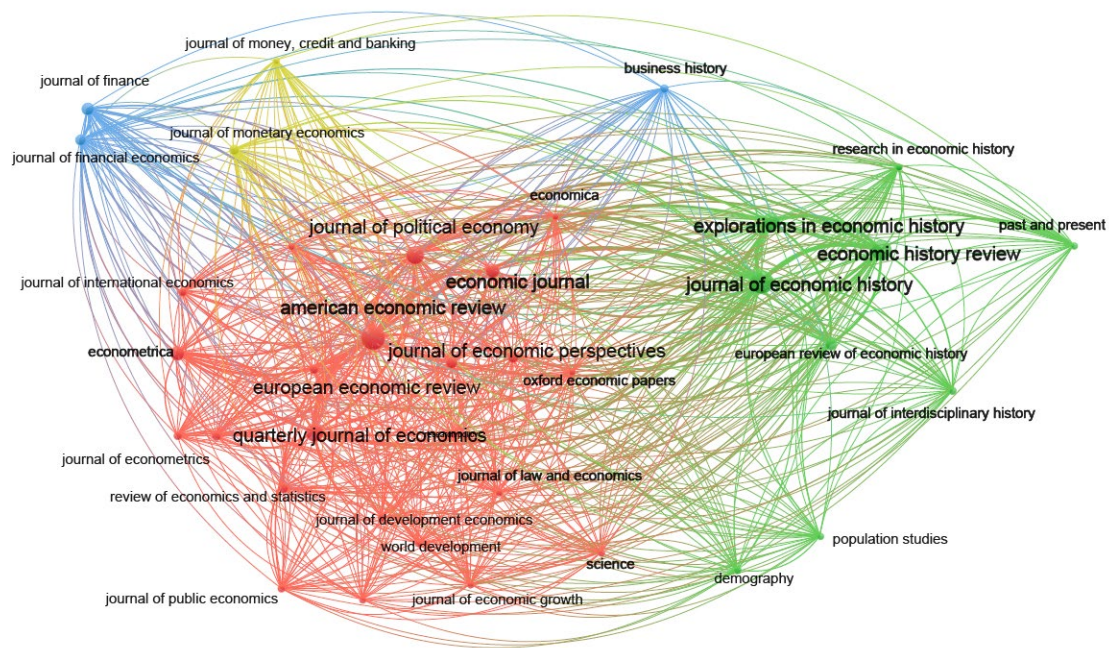


Figure 4. Journals with more than 200 citations in EconHist.

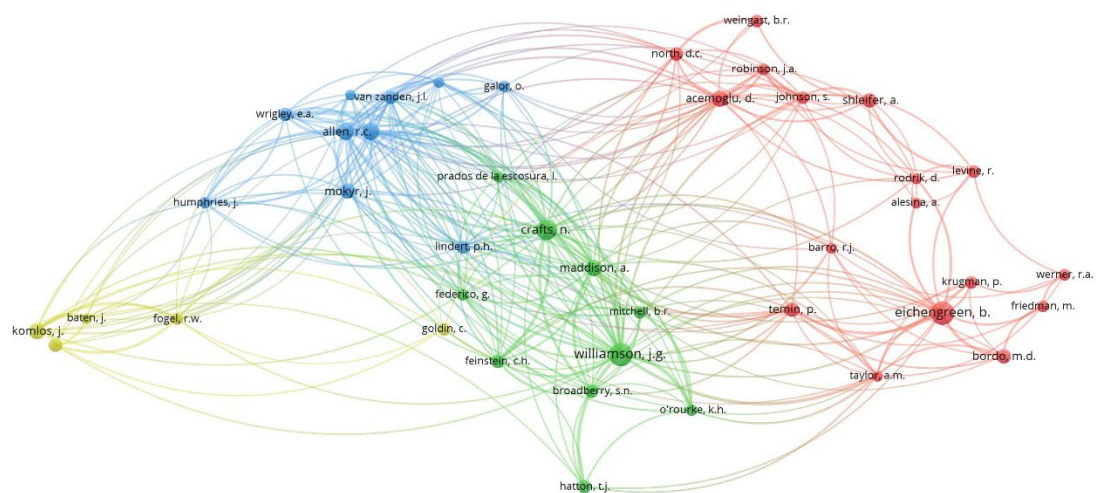


Figure 5. Authors with more than 200 citations in EconHist. Colors refer to citation clusters.

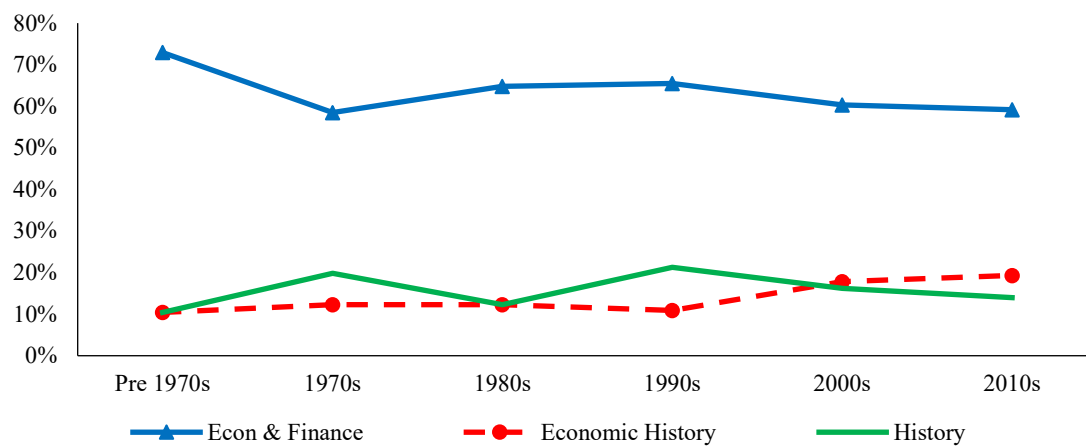


Figure 6. Share of Ph.Ds. by discipline and by Decade of Graduation.

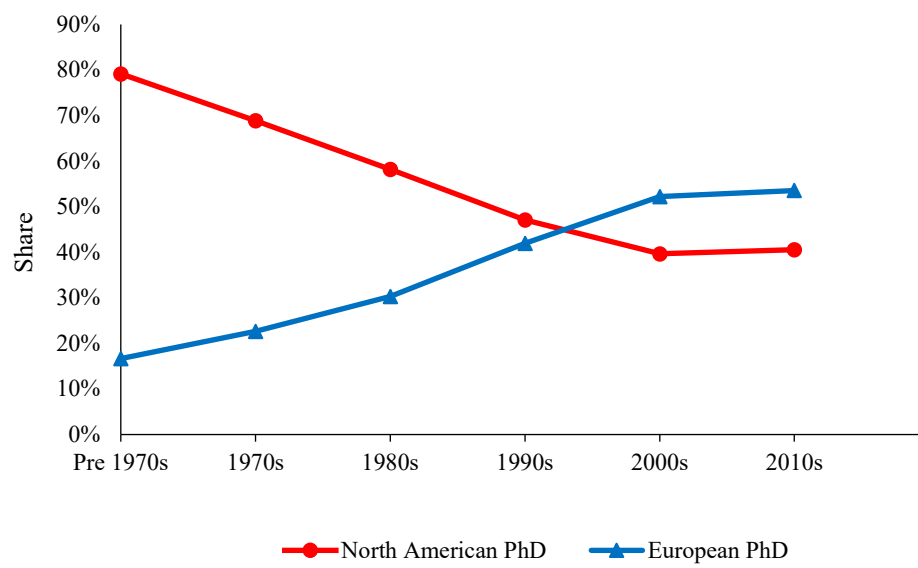


Figure 7. Region of Ph.D. by Decade of Graduation.

Tables

Table 1. Individuals by decade when they were awarded their Ph.D. Degree.

Decade of PhD	Observations	Percentage
1950	6	0.54
1960	43	3.88
1970	107	9.66
1980	121	10.92
1990	172	15.52
2000	249	22.47
2010	410	37.00
Total	1108	100.00

Table 2. Distribution of articles by number of authors.

Number of Authors	N. of articles	Percentage
1	4,259	39.53
2	3,897	36.17
3	1,791	16.62
4	522	4.85
5	232	2.15
6 or more	72	0.67
Total	10,773	100.00

Table 3. A comparison of different databases used in similar papers in the last ten years.

Paper	Selection criteria	# of records
Cioni et al. (2021)	Top five ⁽ⁱ⁾ field economic history journals and thirteen prominent economics journals.	3,500
Galofré-Vilà (2020)	Peer-reviewed articles published in the leading economic history journals since 1980.	5,330
Fourie (2019)	A complete dataset of articles in the top four ⁽ⁱⁱ⁾ economic history journals. It documents the rise in African economic history in the last two decades.	2,293
Margo (2018)	Use of econometric language in articles appearing in academic journals of economic history and economics; and publication histories of successive cohorts of Ph.Ds. in the first decade since receiving the doctorate. ⁽ⁱⁱⁱ⁾	167
Seltzer (2018)	Bibliometric data from the <i>Australian Economic History Review</i> from 1956 through November 2017 to explore research trends at the Journal. Data set is every article published in the <i>AEHR</i> . It adds the citation count collected from <i>Publish or Perish</i> or from <i>Google Scholar</i> whether each article uses econometric analysis, and (if so) the number of observations in the largest regression. It also collects data on co-authorship, citation, and use of econometrics from <i>The Journal of Economic History</i> , <i>Explorations in Economic History</i> , and <i>Business History Review</i> and <i>Business History</i> ; and two ‘regional’ economic history journals: <i>Scandinavian Economic History Review</i> and <i>Revista de Historia Economica</i> .	594 ^(iv)
Wehrheim (2018)	Articles published in <i>The Journal of Economic History</i> between 1941 and 2016.	2,675
Angrist, Azoulay, Ellison, Hill, and Lu (2017)	Dataset 138,892 papers published in 80 journals between 1980 and 2015. Data collected from the WoS and Econlit. ⁴⁷	138,892 (4,850)

⁴⁷ This paper uses machine-learning-based classification of economics journal content into fields and styles, developed as part of a project analyzing citations. The training dataset contains 5,850 papers: 1,507 hand-classified for use in Ellison (2002); and 3,343 additional randomly selected papers hand-classified mostly by our research assistants—thus the number in brackets.

Abramitzky (2015)	Partially based on history papers published with the top journals over the last 45 years. It traces 66 Ph.D. students who graduated between 2010 and 2014 in economic history from the top eight economics departments and from other departments that typically produce economic historians and their 1,032 peers.	N.A. ^(v)
Baten and Muschallik (2012)	Uses Econlit (2005-2010), nine economic history journals and an email questionnaire. 59 countries were surveyed in this overview.	825

Notes: in many cases they are not exactly databases but samples.

(i) *Economic History Review*, *The Journal of Economic History*, *Explorations in Economic History*, *The European Review of Economic History* and *Cliometrica*.

(ii) Fourie excludes *Cliometrica* from the previous top five list.

(iii) It compares Economic history and Labor economics: *The Journal of Economic History* and *Explorations in Economic History* vs. *Industrial and Labor Relations Review*, *Journal of Human Resources* and *The American Economic Review*.

(iv) The author does not provide the number of records; we have inferred it from data in the paper.

(v) The author does not provide the number but the percentage of papers.

Table 4. Doctoral degree obtained by individuals in EconHist database.

Ph.D. Program	# Authors	Percentage
Economics	621	56.66
History	147	13.41
Economic History	142	12.96
Economic and Social History	25	2.28
Finance	17	1.55
Political Science	10	0.91
History and Civilization	8	0.73
Political Economy	6	0.55
Humanities	5	0.46
Law	5	0.46
Management	5	0.46
Social Science	5	0.46
Other	99	9.03
TOTAL	1,096	100.00

Table 5. Probit Average Marginal Effects for Being Awarded a Ph.D. in Economics.

VARIABLES	(1) Econ Ph.D.	(2) Econ Ph.D.	(3) Econ Ph.D.	(4) Econ Ph.D.
Year Ph.D.	-0.002* [0.001]	-0.001 [0.001]	0.002 [0.001]	0.004*** [0.001]
Male		0.085** [0.034]	0.077** [0.031]	0.065** [0.030]
<u>Continent Ph.D.</u>				
Asia			0.184** [0.092]	0.200** [0.088]
North America			0.345*** [0.023]	0.270*** [0.028]
Oceania			0.105 [0.109]	0.127 [0.102]
South America			0.201 [0.240]	0.127 [0.215]
<u>Affiliation</u>				
Cliometric Society (CS)				0.232*** [0.028]
Economic History Association (EHA)				-0.011 [0.029]
European Historical Economics Society (EHES)				0.027 [0.033]
Economic History Society (EHS)				-0.128*** [0.033]
Observations	1,100	1,100	1,098	1,098
Robust standard errors in brackets				
*** p<0.01, ** p<0.05, * p<0.1				