FORSCHUNGSDATEN

TOBIAS A. JOPP / MARK SPOERER

The German Cliometrics Database

ABSTRACT: This short article introduces the German Cliometrics Database as the foundation for an article by Jopp and Spoerer (2024) who trace cliometric research on German history. This newly constructed database of every publication that (1) contributes to the historiography of Germany and (2) employs, as a baseline, inferential statistics enables researchers to find cliometric studies related to their own work much more efficiently. Even though no full texts are provided alongside the data file, the collected abstracts or, respectively, summaries for every publication in the database allow for some baseline text mining approaches. Along with the remaining information provided, they may also form the basis for broader bibliometric or historiographical studies.

Keywords: cliometrics, economic history, Germany, German history, quantitative history, social history

JEL Codes: B4, B16, C01, N01

1. Introduction

In the 1970s, quantitative methods increasingly found their way into German historical studies. On the one hand, this was due to historians' increased interest in economic and social structures and developments, and on the other hand to the easier applicability of statistical methods through the increasing use of microcomputers and corresponding software.[1] However, as we show in our article in the Handbook of Cliometrics[2], in which we specifically trace cliometric research on German history, there was by no means a straight line from historical social science or the Bielefeld School to the increased use of quantitative methods, often referred to as cliometrics. The early cliometric works on German history in the initial period, which we date to the years 1977 to 1991, dealt to a large extent with social and politico-historical topics such as demography, crime and voting behavior and were written by historians with an interest in quantitative methods. In the 1990s, this historical root of cliometrics was joined by an economic root, which

---

[1] For a recent study tracing the business cycle of quantitative methods in German historiography, see Buchner et al. (2020). Ruggles and Magnuson (2019) provide a methodically comparable, yet geographically broader study.

[2] See Jopp and Spoerer (2024).

soon became dominant and has remained so to date. Authors in this field had studied economics and, as part of this, had also undergone solid statistical training, making it easier for them than the historians to adopt the increasingly differentiated canon of econometric methods. Their research interest naturally focused on genuinely economic-historical questions, so that today cliometrics is often understood as the quantitative part of economic historiography. However, we believe that this equation is too simplistic, precisely because early cliometric work shows that quantitative methods can also be used very fruitfully to address non-economic-historical issues.

Our knowledge history of cliometric historical research, if you will, is based on the quantitative evaluation of a database of over 800 publications that have been collected specifically for this purpose. The publications meet two conditions: Firstly, they deal with German history (as researched by authors institutionally affiliated in- as well as outside Germany); and, secondly, they use advanced statistical methods. Along with this article, we are publishing our "German Cliometrics Database" so that other researchers can continue to work with it and help supplement it if necessary.

Formally, we think our database can serve two main purposes. Firstly, it helps researchers find cliometric studies relating to a specific topic of German history much quicker than by searching in other ways. Secondly, and more importantly, since we analyzed the database using only very simple methods of descriptive statistics to, for example, highlight certain trends in publication behavior, the regional distribution of involved research facilities, and the business cycle of research topics, many more questions may be answered with the assembled (bibliometric) data in combination with appropriate methods. To give just a few thought-provoking impulses: From a gender-historical perspective, it may be interesting to see whether publications by female authors, which have been coming at a numerical disadvantage in quantitative and, specifically, economic history, systematically focus on different topics than publications by male authors. Or, since the dataset provides all authors and contains information on their affiliation, the increased internationalization of German historians can be studied in greater detail, thereby identifying research(er) networks and their publication success.

The greatest potential for insight into the history of the discipline would probably come from a comparison of the summary characteristics of the corpus, which we refer to as cliometric and present here, with the much larger group of all publications on German history, that is, all remaining quantitative and the many more qualitative publications. The universe of relevant publications would not necessarily have to be compiled individually as here – the order of magnitude is probably in the six-figure range (at least according to ChatGPT) – but might be taken from bibliographic databases. In our handbook article and previous work, we approached the issue of the true proportion of quantitative historiographical studies only very crudely.

This article proceeds by explaining the search and selection mechanism leading to a publication's inclusion in the database (Section 2), introducing the structure of the data file (Section 3), and providing a short descriptive overview of the corpus's basic features (Section 4). The article ends with a short conclusion.

2. Remarks on the data gathering process

We made the following decisions on what would constitute "Germany", which applied methods would qualify as "cliometric", and which types of publication to include.[3] As for the geographical aspect, we decided to focus on regions in Central Europe where the German language was dominant, but explicitly excluding Austria and Switzerland in today's borders.[4] Hence, the Germany we refer to for the purpose of selecting suitable publications may best be imagined as Germany in the borders of 1937. Note that we not only included studies dealing exclusively with German history but also those which deal with Germany as one of several European or world regions. In the latter case, however, we decided to include only those studies that provide explicit historiographical discussions of the empirical findings on Germany.[5]

Furthermore, we defined a study to be cliometric if, as a baseline, it applied inferential statistics to historical data on Germany, beginning with showing significance values for descriptive measures like correlation coefficients.[6] To hard core cliometricians, this may seem like a quite low entry threshold, but it was important to us to include (or "honor") early quantitative work in the corpus. By using this definition, we do put the focus on the tip of the iceberg of quantitative historiographical studies, if you will, and unavoidably exclude quite many of them that would be empirical, but purely (and, possibly, extensively) descriptive in nature. This decision was made because we wanted to put emphasis on studies putting forward causal arguments in a formal way; besides, we thereby also kept the effort manageable. However, we made an exception to this selection rule in that we additionally included studies applying non-trivial mathematical methods like growth accounting and productivity decomposition, network analysis, or text mining approaches based on probabilistic algorithms.

As for the types of publication, we decided to not narrow down the focus on journal articles, but consider also monographs, book chapters, and the most recent working papers (that is, those published in 2020 or later). However, we excluded unpublished cumulative dissertations if the papers they consist of would appear as a journal article (but we kept non-cumulative ones).[7]

Finally, some remarks on our search strategy or, respectively, efforts are needed. For all steps we briefly go through in the following it holds that we collected and manually checked every publication that seemed to be a potential candidate for our database – be it because the publication's title suggested so or because an available abstract, cover page, or summary did; when in doubt, we rather checked than neglected the publica-

---

3   The focus on works dealing with German history itself was predetermined by the editors of the handbook.
4   This also ruled out Alsace-Lorraine and regions in the East where the predominant language was, for example, Polish.
5   In other words, it is not sufficient to derive empirical results partly from data on Germany; the results' implications for German history must be explicitly discussed.
6   For the definition of historical data, implicitly used by us, see Jopp and Spoerer (2023).
7   An example of the latter would be the unpublished dissertation of Kirchhain (1973).

tion. It goes without saying that we skimmed through far more publications than actually made it into our database.

In a first step, we manually went through all issues of historiographical journal outlets – economic-historical and otherwise – we suspected relevant quantitative studies to have been appearing in on a more regular basis.[8] Informed by some chance finds occurring in later steps, we also did so with some German-language economics journals.[9] Additionally, we performed keyword searches directly on the homepages of many remaining economics and also several miscellaneous journals that we suspected of yielding potentially relevant contributions.[10] In a second step, we screened relevant book series, and in a third step, we went, if available, through the CVs or, respectively, list of publications of all (economic) historians who came to our mind based on the preceding steps and checked up on each publication's relevancy; the latter was especially helpful in identifying a core of collective volume articles and some journal articles in rather exotic outlets. Only in a fourth step did we turn to established bibliographic databases like JSTOR, EconLit, and Historical Abstracts and performed another round of keyword and Boolean searches to identify yet still missing publications. Finally, it is worth noting that we did not systematically browse through the reference lists of all publications we pre-selected for checking; the reason being time constraints.

Even though we cannot be sure to have identified virtually every publication on German history meeting our criteria, we are confident that, with our search strategy, we have been coming up with a reasonably good initial guess.

3. Organization of the data file

The provided data file consists of two spreadsheets, with the first one containing the actual data on the collected publications and the second one reporting a list of journal abbreviations used for the sake of convenience. The data spreadsheet's first row contains the labels on the variables gathered or, respectively, measured per publication. The columns are organized as follows:

---

8   E. g., the international top five economic history journals (Cliometrica, Economic History Review, European Review of Economic History, Explorations in Economic History, Journal of Economic History) as well as the national top three (Jahrbuch für Wirtschaftsgeschichte, Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte, Zeitschrift für Unternehmensgeschichte); journals on quantitative history (including but not limited to the field of economic history), namely Historical Methods, Historical Social Research, Journal of Interdisciplinary History, and Social Science History; and many more national and international economic, business, social, and miscellaneous history outlets like Archiv für Sozialgeschichte, Business History, Business History Review, Central European History, Continuity and Change, Geschichte und Gesellschaft, Historische Zeitschrift, and so on.
9   E. g., Jahrbücher für Nationalökonomie und Statistik, Zeitschrift für die gesamte Staatswissenschaft (Journal of Institutional and Theoretical Economics), and Weltwirtschaftliches Archiv.
10 Among these were journals such as American Economic Review, American Political Science Review, Demography, Journal of Economic Growth, Journal of Political Economy, Population Studies, and Quarterly Journal of Economics.

- *Column 1*: Contains the entry number (running from 1 to *N*, with *N* being the total number of publications, namely 816).
- *Column 2*: Identifies all authors of a publication by last and first names, with the authors separated by semicolon. For each publication, authors are given in the order they appear in the publication, which in most cases is alphabetical. The sorting order of the publications is chronological and within a year alphabetical by the first given author's last name.
- *Column 3*: States the publication's year of publication.
- *Column 4*: States the publication's full title in the original publication language (we leave it to the users to implement suitable translations in case they are needed).
- *Column 5*: Provides summaries of the publication's data basis, methodical approach, and main results. In most cases, this summary is the publication's abstract itself, to be found either in the text file we secured or on the respective journal's or publisher's homepage (or someplace else on the internet). In quite a few instances (of mostly older publications) though, namely 156, there was no abstract available, whatsoever. In such cases, we extracted a suitable text part from a publication's introduction or conclusion to fill the gap. This handling is indicated by an expression in square brackets at the beginning of the summary, stating from where the text part was taken (e. g., "[Taken from introduction]"). As with the title, the summary is given in the original publication language.
- *Column 6*: Gives a publication's place of publication. For journal articles, this is the journal name, abbreviated according to the list reported in spreadsheet two (e. g., "JIH" for "Journal of Interdisciplinary History").[11] For collective volume articles, chapters, and monographs, in turn, this is the publisher and its geographical location. For (most recent) working papers, this is the institution providing the publication platform (e. g., "CEPR").
- *Column 7*: Identifies the type of publication, that is, "chapter", "journal article", "monograph", and "working paper".
- *Column 8*: For journal articles, identifies the field of journal, that is, "econhist" (economic history), "(Remaining) history", "economics", and "other"; otherwise, it reads "none".
- *Column 9*: Identifies publications written in English by a dummy taking the value of 1 (0 otherwise).
- *Column 10*: Identifies publications written by a single author by a dummy taking the value of 1 (0 otherwise).
- *Column 11*: Identifies publications exclusively authored or co-authored by (a) female author(s) by a dummy taking the value of 1 (0 otherwise).
- *Column 12*: In the order of appearance as per column 2, provides the exact institutional affiliation of each author as reported in the publication. In case more than one affiliation per author is reported in the publication, we collected the first affiliation

---

11   The chosen abbreviations are to our own liking, thus may not fully conform with officially used abbreviations.

mentioned, assuming that this is the most important one. We only made an exception to this procedure if the first reported affiliation was to an institution like the CEPR, to which especially many economists have been research fellows on top of one or more university affiliations. We then collected the first mentioned university affiliation instead.

– *Column 13*: In case we were able to identify, we add the organizational unit(s) of the author(s) here (again in the order of appearance as per column 2); otherwise, it reads "none".

– *Column 14*: Gives the country in which the institution as per column 12 is located.

– *Column 15*: Identifies publications exclusively written by authors affiliated outside Germany by a dummy taking the value of 1 (0 otherwise).

– *Column 16*: Gives the geographical focus of the publication as per the used data basis; either Germany in full (i. e., "Germany") or a particular German region or regions (e. g., "Baden" or "Prussia").

– *Column 17*: Identifies the studied historical epoch(s) which we defined and abbreviated as "Middle Ages/MA" (<1500), "Early Modern/EM" (1500–1800), "post Vienna Conference Germany/PV" (1800–1870), "Empire without WW I/EMP" (1871–1914), "World War I/WWI" (1914–1918), "Weimar Republic/WR" (1919–1932), "Nazi period without WWII/NZ" (1933–1938), "World War II/WWII" (1939–1945), "Post-WWII-West/PWW" (>1945), and Post-WWII-East/PWE (1946–1989).

– *Column 18*: Based on the publication's epochal focus as per column 17, identifies a publication as a long-term study by a dummy taking the value of 1 (0 otherwise). To qualify as a publication with a long-term focus, a publication had to meet one of the following two conditions: (1) it shows the epochal focus "MA-EM", "MA-EM-PV", "EM-PV" or "PV-EMP"; or (2) it spans at least five epochs whatsoever, a definition helping to avoid counting the combination "WW1-WR-NZ-WW2" which makes for too brief a study period in our view.

– *Column 19*: Identifies the main, or meta, topic of the publication as per our definition (see Section 4 for the abbreviations used). We defined 20 meta topics (see Table 2 below for an overview), and each publication was assigned exactly one such topic. The defining work and assignment procedure was entirely done by hand (or, put differently, by "close reading"), meaning, without support by a (seemingly) objective algorithm like topic modelling allowing for "distant reading". We arrived at the number of 20 meta topics by repeatedly grouping the publications according to sub-topics (see the following column) until we found a seemingly stable distribution. The number of twenty therefore is a chance product.[12]

– *Column 20*: Meant to provide some differentiation vis-à-vis the assigned meta topic, additional sub-topics are provided for each publication; assignment was again per hand.

---

12 Wehrheim et al. (2022) is an example for an approach supported by such an algorithm. However, we saw some methodical and technical issues with our dataset (e. g., different types of publications meaning different text lengths) that let us to abstain from a topic modelling approach to arrive at meta topics.

Note again that we do not provide full texts along with the data file in the ZBW Journal Data Archive by default, but abstracts/summaries only. However, we do have full texts on each entry in our database at hand.

4. Data description

As of March 2023, when we stopped including studies for the time being, our database comprises *N*=816 publications – unsurprisingly, most of them journal articles and linked with the subfield of economic history. Table 1 provides a basic overview of our corpus's features; for a more detailed look at trends in the data, we refer readers to our handbook article.

Among the 816 publications in our database we labeled "cliometric" are 686 journal articles, 43 chapters or, respectively, collective volume articles, 66 monographs, and 21 working papers, with the very first two studies written by Greek economist John S. Pesmazoglu in 1950 and, respectively, 1951; both are journal articles published in English in two leading German economics journals of the time, namely *Weltwirtschaftliches Archiv* and *Zeitschrift für die gesamte Staatswissenschaft*.[13] We found 252 more articles in economics journals as well as 281 in economic history outlets, 99 in remaining history journals, and 52 in journals relating to disciplines other than history or economics. In total, almost four fifths of the publications have been written in English; as might be expected, the percentage is significantly higher for journal articles and recent working papers than for the other two types of publication.

Furthermore, only about eight percent of the publications have been exclusively authored by female authors, with an additional ten percent at least co-authored by female authors; it is not an exaggeration therefore to say that the subfield of cliometric research on German history is dominated by male authors and, potentially, "male subjects". However, whether there are typical "male research subjects", and therefore also "female research subjects", and whether the unequal distribution of sexes among authors is representative for the entirety of cliometric studies on any theme anywhere in the world awaits further investigation.

Table 1 also shows that no less than almost 45 percent of the publications were written by authors who were, at the time of finalizing the publication, affiliated to research facilities outside Germany.[14] Looking into the type of publication, this percentage is even higher for articles published in economics and miscellaneous field journals.

As for the regional focus, 22 percent of the publications have a decidedly regional focus, in many cases on Prussia which, to a considerable part, might be explainable by data availability. In turn, roughly four fifths of the collected studies mean to directly derive results for Germany as a whole. As for the time periods studied, only seven percent of

---

13   See Pesmazoglu (1950, 1951).
14   Note that this observation does, of course, not imply anything on the authors' nationality, which we did not collect.

Table 1: Basic corpus characteristics

Type of publication	Time coverage	#	Written in English	Fully male-authored	Fully female-authored	Authors fully affiliated outside Germany	Focus on specific German regions	Focus entirely on premodern period	Long-term study
Chapters	1956–2021	43	25 (58.1%)	35 (81.4%)	4 (9.3%)	14 (32.5%)	12 (28.0%)	5 (11.6%)	6 (13.9%)
Journal articles	1950–2023	686	579 (84.4%)	564 (82.2%)	50 (7.3%)	329 (47.9%)	149 (21.7%)	44 (6.4%)	114 (16.6%)
Economics journals	1950–2023	254	227 (89.4%)	213 (83.8%)	7 (2.7%)	133 (52.4%)	40 (15.7%)	9 (3.5%)	46 (18.1%)
Economic history journals	1973–2023	281	245 (87.2%)	224 (79.7%)	32 (11.4%)	122 (43.4%)	78 (27.7%)	20 (7.1%)	37 (13.2%)
History journals	1973–2023	99	64 (64.6%)	82 (82.8%)	8 (8.1%)	39 (39.4%)	29 (29.3%)	11 (11.1%)	19 (19.2%)
Miscellaneous journals	1976–2022	52	43 (82.7%)	45 (86.5%)	3 (5.8%)	35 (67.3%)	12 (23.1%)	4 (7.7%)	12 (23.1%)
Monographs	1962–2022	66	20 (30.3%)	54 (81.8%)	11 (16.7%)	16 (24.2%)	15 (22.7%)	5 (7.6%)	11 (16.7%)
Working papers	2020–2022	21	21 (100.0%)	14 (66.7%)	2 (9.5%)	6 (28.6%)	5 (23.8%)	3 (14.3%)	6 (28.6%)
Total	1950–2023	816	645 (79.0%)	667 (81.7%)	67 (8.2%)	365 (44.7%)	181 (22.2%)	57 (7.0%)	137 (16.8%)

Notes: Percentages given for the various types of publication refer to the types of publication subtotal; percentages in the last row to the grand total.
Source: Authors' own depiction.

the publications have a focus on the premodern period (as opposed to the 19th and 20th centuries), which may likewise be explained to a large part by data availability issues that make advanced quantitative studies impossible. However, as the percentage for working papers, which are not older than 2020, shows, studying the premodern period cliometrically is slowly attracting more and more interest, as, generally, is conducting truly long-term studies spanning at least around a century of data.

Finally, Table 2 summarizes the meta topics we assigned to each publication. Given are the topics in full, the abbreviations we use in the data file, and the number as well as percentage share of publications subsuming under the meta topics. A third of all publications in the database relate to three top meta topics, namely to "economic growth (GDP)", "demography (DEM)", and "sectors (SEC)". Unsurprising to many readers may come the fact that the topic having yet produced the least number of cliometric publications is "culture (CUL)". However, we like to see the positive message in this observation, which is that cultural history can, indeed, be a subject of cliometric research.

Table 2: Publications by meta topic

Meta topic	Short-hand	# Publica-tions	Share
Culture (incl. private consumption)	CUL	8	1.0 %
Currency (incl. gold standard, Bretton Woods etc.)	CUR	27	3.3 %
Demography (incl. migration, forced displacement)	DEM	92	11.3 %
Banking and finance (incl. capital markets, stock exchange etc.)	FIN	80	9.8 %
Economic growth (including productivity)	GDP	96	11.8 %
German Democratic Republic	GDR	17	2.1 %
(Economic) Geography	GEO	23	2.8 %
Human capital (incl. education)	HUM	32	3.9 %
Income (incl. wealth and anthropometrics)	INC	60	7.3 %
Institutions	INS	22	2.7 %
Labour	LAB	11	1.3 %
National Socialism (incl. Nazi voters 1930–32)	NAZ	58	7.1 %
Politics (incl. elections, jurisdiction, political parties, protest)	POL	28	3.4 %
Prices (incl. wages, market integration)	PRI	71	8.7 %
Public finances	PUB	16	2.0 %
Religion (incl. antisemitism)	REL	15	1.8 %
Sectors (incl. agriculture, firms, railroads, concentration)	SEC	86	10.5 %
Social history (incl. social security, social mobility, crime)	SOC	29	3.5 %
Technology (incl. patents)	TEC	30	3.7 %
Trade (incl. foreign trade, globalization, tariffs)	TRA	15	1.8 %
20 meta topics		816	100.0 %

Source: Replicated from Jopp/Spoerer (2024), p. 23.

5. Concluding remarks

This database is open in the sense that we plan on updating it regularly – by including research yet to come as well by past publications we have not found yet. Thus, the number of entries will be further rising beyond its current state, and the database's layout may also be subject to adjustment in the future. To accomplish the aim of identifying every relevant cliometric study on German history (according to our definition on what a cliometric study is), we want to explicitly encourage readers to bring every past publication to our attention that we might have missed out on during our search. We will thoroughly examine every proposed study for possible inclusion in our database.

Dataset

DOI: 10.15456/vswg.2024078.1048204229

References

Buchner, Michael / Jopp, Tobias A. / Spoerer, Mark / Wehrheim, Lino (2020): Zur Konjunktur des Zählens – oder wie man Quantifizierung quantifiziert. Eine empirische Analyse der Anwendung quantitativer Methoden in der deutschen Geschichtswissenschaft, in: *Historische Zeitschrift* 310, pp. 580-621.

Jopp, Tobias A. / Spoerer, Mark (2023): Historisch-statistische Daten, in: *Silvia Daniel et al. (eds.): Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, 3rd edn., Berlin, pp. B.10-1-B.10-31 (https://guides.clio-online.de/guides/sammlungen/historisch-statistische-daten/2023).*

Jopp, Tobias A. / Spoerer, Mark (2024): Cliometrics and the study of German history, in: Claude Diebolt / Michael Haupert (eds.): *Handbook of Cliometrics*, 3rd edn, Berlin/Heidelberg, pp. 1-54.

Kirchhain, Günter (1973): *Das Wachstum der deutschen Baumwollindustrie im 19. Jahrhundert. Eine historische Modellstudie zur empirischen Wachstumsforschung*, unpubl. Ph.D. diss., Münster.

Pesmazoglu, John S. (1950): Some international aspects of German cyclical fluctuations, 1880-1913, in: *Weltwirtschaftliches Archiv* 64, pp. 75-110.

Pesmazoglu, John S. (1951): A note on the cyclical fluctuations of the volume of German home investment, 1880–1913, in: *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics* 107, pp. 151-171.

Ruggles, Steven / Magnuson, Diana L. (2020): The history of quantification in history: the JIH as a case study, in: *Journal of Interdisciplinary History* 50, pp. 363-381,

Wehrheim, Lino / Spoerer, Mark / Jopp, Tobias A. (2022): Turn, turn, turn: A digital history of German historiography, 1950-2019, in: *Journal of Interdisciplinary History* 53, pp. 471-507.

TOBIAS A. JOPP
Universität Regensburg
Lehrstuhl für Wirtschafts- und Sozialgeschichte
D-93040 Regensburg
tobias.jopp@geschichte.uni-regensburg.de

MARK SPOERER
Universität Regensburg
Lehrstuhl für Wirtschafts- und Sozialgeschichte
D-93040 Regensburg
mark.spoerer@geschichte.uni-regensburg.de