

Back to the Sources

Claire Lemercier, Claire Zalc

▶ **To cite this version:**

Claire Lemercier, Claire Zalc. Back to the Sources: Practicing and Teaching Quantitative History in the 2020s. 2020. hal-03037343v1

HAL Id: hal-03037343

https://sciencespo.hal.science/hal-03037343v1

Preprint submitted on 3 Dec 2020 (v1), last revised 15 Sep 2021 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Version préliminaire d'un article pour la revue *Capitalism*.

Back to the Sources
Practicing and Teaching Quantitative History in the 2020s

Claire Lemercier, CNRS, Paris and Claire Zalc, CNRS & EHESS, Paris

This paper elaborates on our joint experience of teaching quantitative methods to (mostly) historians since the early 2000s and writing an introductory book on this topic, first in French, then in English, in a revised and expanded version.[1] All along, we have pursued three aims, related to two different types of audience.[2]

First, we want to make quantitative methods accessible for all historians—and humanists generally—, especially those who do not think that such methods are "for them," because they do not enjoy mathematics, or because they study topics that are not traditionally considered as suited to quantification. We have sometimes written that we wanted to make quantitative history banal, in the sense that it would be published in regular historical journals without being remarked upon for being quantitative.

Second, our intent is to contribute to less routine uses of quantification in the social sciences, by promoting diversity in methods and imagination in categorization schemes—going beyond "the usual suspects" in terms of sources, variables, and calculations. Our intended audience here includes seasoned practitioners of quantification.

Third, we promote respect for the basic tenets of the historical profession, i.e. principles of source criticism, as the cornerstone of the constitution of data from historical sources. This third goal has become more and more central for us—hence its presence in the title of this paper. It is a message that we want to convey not only to non-historians, but also to historians who, because they are not used to quantification, tend to equate it with an automatic reading and interpretation of the source by a computer, or with the imposition of simplistic categories.

We are aware of the risk of trying to make the most of the insights of quantitative social science *and* critical humanities: we might end up convincing neither of the two audiences[3]. We nevertheless try to make the best out of our bridging position. It is easy to translate in the context of this journal. New historians of capitalism who apply methods from cultural history to economic topics are close to our "average historian" audience (as are most business historians, whose bread and butter is narrative case studies). As their approach is in part a political and methodological reaction to cliometrics, many tend to link any type of quantification or measurement with neoclassical economics and commodification.[4] Conversely, few economic historians in economics departments will think that they need advice from self-trained historians whose book mostly discusses examples in social, cultural, and political history, does not give formulas or address standard deviation, but devotes many pages to sequence analysis and network analysis. In terms of careers, young researchers in the two groups probably have no

1 Lemercier & Zalc, *Méthodes quantitatives* ; Lemercier & Zalc, *Quantitative Methods*, with a companion website: https://quanthum.hypotheses.org/

2 We have written our books with the two audiences in mind, which explains their non-standard format. See Lemercier, "How should I read this book?"

3 Sociologist Andrew D. Abbott discussed it in *Time Matters*, 281.

4 As do other critics of standard economic history, e.g. Boldizzoni, *The Poverty of Clio*.

incentive to learn non-standard quantification. We however hope that, in each of the two groups, some colleagues are already fond enough of sources and experiments to hear us.[5]

This hope is grounded in our teaching experience with graduate students. Not only did we convince many of them, over the last twenty years, that our approach could be useful. In fact, it is this teaching, and not only our personal research experience, that led us to make our stance on quantification more consistent and explicit over the years—as we saw what worked and did not work with diverse actual sources and research questions. This paper aims at summing up what we learned from this experience. Our books are self-help handbooks: any person beginning a research can use them to learn the basics on their own, because we know that few will be able to attend a course. We hope that this paper will help colleagues in history who do not think of themselves as quantifiers to embed quantitative history in their methods courses, and that it will help economists who teach economic history to discuss source criticism in the context of quantification.

The first part of the paper begins by explaining where we speak from. As practices of quantification differ between countries and sub-disciplines, we first tell a few words about our own experience with quantitative history, in the context of its recent evolutions, since it lost any pretense at dominance in the historical discipline. These trajectories led us to promote constructivist, small-scale, experimental quantitative history.[6] In terms of teaching, this translates into a learning-by-doing focused on the construction and categorization of data from sources. The second and third parts of the paper briefly flesh out the main principles that we promote in our teaching, with examples in and out of economic history and the history of capitalism. The second part addresses the transformation of sources into quantifiable data; the third part discusses data categorization and analysis.

1. From Self-Teaching to Having Others Learn by Doing

We came to quantitative methods more out of necessity than out of faith: our historical materials came in too large quantities to be manageable with just close reading and .doc notes. This "why" may help to understand how we teach and practice quantitative history: we give the courses that we would have wanted to follow when we were beginners. This implies that the courses should be open to all graduate students (and more advanced colleagues who want to learn) and should avoid fascination for the tools themselves, as opposed to the intellectual process of quantification. Our teaching aims at changing the practices of the quantifiers as well as the average historian.

1.1 Teaching Quantitative History—But which one?

We did not begin our research, in the late 1990s, thinking that we would become quantitative historians. Claire Zalc's master's dissertation on German and Austrian migrants in France in the 1930s was entirely based upon discursive sources. But for her PhD dissertation, which focused on immigrant shopkeepers in the interwar period, her main source, the Business

5 For similar positions, see Lamoreaux, "The Future of Economic", and Rosenthal, "Seeking a Quantitative Middle Ground".

6 In addition to our book, see Karila-Cohen et al., "Quantitative History." The fact that the co-editors are four women is not anecdotal for us. Freeing quantitative history from its gendered connotations is part of our more general effort to de-standardize it.

register, was massive (over 1 million inscriptions).[7] Quantification was not an end in itself or a religion, but one tool among others, necessary to sample this massive source and use it to write better history. Just before she began her PhD, Claire Lemercier learned the basics of Excel (mundane but crucial functions like "Freeze panes" or "Pivot table") from a staff member in a historical demography research center. She gathered and analyzed data on the institutional careers and personal ties of a few hundred Parisians in the nineteenth century, but it was just one part of her dissertation on the advisory role of the Chamber of Commerce. And for her, if was not "quantitative history." Neither of us had any reason to identify with a label that was definitely old school: in the generation of our PhD advisers, many had learned how to quantify, some had done it, but almost none still thought that it was relevant.

In our view, our story tells a lot about the position of quantification in history since the early 2000s. Our chance was to know social scientists who shared tools with us, but it was not easy for us to find methods courses, or history professors able to advise us; it was almost impossible for others. Before 2001 and a first handbook presenting the basic use of Excel and Access, which stimulated our own,[8] the main book on quantitative history available in French addressed the use of a calculator to compute moving averages or variance on aggregate economic or demographic data. Yet we were perhaps lucky to arrive at a moment when the war against the first wave of quantification had been won: many historians still equated quantification with anachronism, structure ignoring agency, or just boredom, but indifference was more widespread than attacks.

The story of this first wave has been told elsewhere.[9] We do not want to retell it here, but we insist on the fact that it should not be forgotten. On the one hand, it produced some excellent research that we like our students to read. On the other hand, it also receded for important reasons that we need to keep in mind if we do not want to reproduce past mistakes. What we call "the first wave" here is the "new economic history" (also called cliometrics), "new social history," and "new political history" that were so branded around the 1960s. Two reasons for disillusionment with this first wave remain central for us. First, some of its promoters had sometimes dispensed with source criticism, imposing anachronistic analytic categories or treating numbers found in past sources as objective data. If historians are to quantify again, they need to collectively learn to adapt source criticism to data construction, and vice versa. "[S]quads of diligent assistants assemble data, encode it, programme it, and pass it through the maw of the computer, all under the autocratic direction of a team-leader."[10] Coming after the first wave had receded, we did not experience this as students. We learned quantification as an individual craft, working on our personal computers, and we came to appreciate this version where the close reading of the source, the construction of data, the quantitative analysis and the interpretation all happened in the same head. We do not reject collective research generally, of course, but we don't want quantification to be equated, again, with the subcontracting of source criticism to assistants.

While we were pursuing our own trajectories, whatever happened to the use of numbers and computers to deal with historical sources? It had not completely disappeared in the 1980s. Our own approach evolved through readings and conversations with four different streams of

---

[7] Zalc, *Immigrants et indépendants*.

8 Cellier & Cocaud, *Traiter des données historiques*.

9 Even though the timing and the main tenets of the first wave (as summed up by e.g. Sewell, "The political unconscious") were commensurate in most countries, there were many differences in topics and methods. We give our own telling in Lemercier & Zalc, *Quantitative Methods*, chapter 1.

10 Stone, "The Revival of Narrative", 6.

research (mostly unconnected to one another). The texts that we have our students read come from all four groups—as well as from the "first wave" and more idiosyncratic corners of the academia.

First, cliometrics mostly evolved in isolation from historians, but in close relation to econometrics, micro-economics, and neo-institutionalist economics. Its mainstream approach in the 2010s involves multivariate regression as a tool aimed at disentangling causes from effects (whereas cliometrics in the 1960s often used quantification in a more descriptive way). Moreover, definitions of causation in econometrics have become more and more stringent, leading many economic historians to focus on specific historical situations, deemed "natural experiments" (where an important event occurred in a country or region, but not in another otherwise similar one).[11] As cliometrics is almost only practiced in economics departments and published in economics journals, it follows the standards of the discipline, which reward other qualities more than source criticism.

Second, "digital humanities" became a popular phrase in the mid-2000s, as did "big data" in the 2010s. Suddenly, there were new series of handbooks and workshops addressing topics that overlapped with ours (network analysis, text analysis) and we were invited to give our opinion, or even thought of as part of the new group. However, this group was indifferent to a lot of what we cared about (the social sciences, contingency tables, longitudinal data). And many of its members came back to what, in our view, had failed in the first wave. We regularly find the same mottos: "retooling" history so that it functions more like the "hard sciences," with more money, more teamwork, and more objectivity; turning the historian into a programmer. The *History Manifesto* went as far as to criticize the undue interest of many historians in archives.[12] It is striking that many of the criticisms made to the "big science" of the first wave ca. 1980 (lavish funding, mathematical sophistication, a quest for exhaustive data, for comparatively few new insights about the past) could be applied without any change.[13] Those who marvel, for example, at Google NGrams without pausing to wonder what exactly is included in Google Books are, for us, not very different from those who admired figures of deflated wages without questioning their sources. And the search for general, unambiguous "ontologies" that would allow machines to read, interpret, and compare all historical texts is, for us, as risky and pointless as were past anachronistic occupational classifications. The good news is that there are exceptions to this mainstream version inside digital humanities[14]—as there are inside cliometrics, of course.

Third, as continental Europeans, we were exposed early on to Italian (social) micro-history. Contrary to most English-speaking versions of micro-history, its criticisms of the first wave did not imply that all quantification was to avoid. The Italians were not postmodern researchers; they did not turn away from the question of standards of evidence[15]. Their answers allowed us to think of quantification as possibly constructivist, suited to the scale of interactions between individuals, and experimental, in the sense of non-standard, non-boring. Meaningful results can be obtained by systematically and even quantitatively studying a small, situated unit (a family,

[11] Due to space limitations, we of course can only provide very simple sketches of each approach here. For mainstream cliometrics, see e.g. Bourgeois-Gironde & Monnet, "Natural Experiments."

12 Armitage & Guldi, *The History Manifesto*. For critical comments, see "Debating the *Longue Durée*."

[13] We find re-reading Stone, "The Revival of Narrative," quite healthy in this regard.

14 For example Weingart, "Punched-Card Humanities", Putnam, "The Transnational and the Text-Searchable".

15 Rosental, "Construire le 'macro' par le 'micro', Trivellato, "Is There a Future," Trivellato, "Microstoria."

a village, etc.). Quantification is not an aim in itself, but it can open new questions or reveal exceptions suited for a more narrative approach. As micro-historians advocate "incorporating into the main body of the narrative the procedures of research itself,"[16] source criticism and data construction may become an integral part of quantitative research—not preliminary operations to be dispensed with or standardized. And this all can be fun. It can be "experimental history," in the sense of trying weird, idiosyncratic categorizations or correlations, in the hope of finding something new[17]: the opposite of the "big science" of the first wave.

Fourth, sociologists and political scientists went on quantifying from historical sources when most historians had stopped—as did economists.[18] We drew a lot of inspiration from their wide set of tools (wider than that of economists).[19] For example, writings by sociologist Andrew Abbott were exciting for us when he called for unheard-of alliances of methods, questions, and materials and criticized the routine use of variables and regression in mainstream US sociology.[20] He also discussed ways to deal with longitudinal data: not only do individual variables change over time, but their meanings and their possible effects also change between historical contexts; sometimes, it seems meaningless to disentangle one effect from a series of causes. Yet the solution is not purely narrative: all this complexity can be modeled, formalized, and visualized.[21]

While we were discovering this motley world of quantification and thinking of ways to introduce it to our students, we decided that our own quantitative history would be constructivist, small-scale, and experimental. Many colleagues had already more or less independently invented and re-invented this practice; we circulate their works as much as we can.[22]

1.2 Teaching: A Focus on Sources Rather than Equations

We have now been teaching quantitative methods for almost twenty years, often together, in different universities in Paris and on occasion elsewhere.[23] We have practiced two main formats. First, in compulsory courses for beginning graduate students in history ("master 1" in the European system), we have the group read some papers and, more importantly, experience collective research, from the close reading of a source to short papers. Second, in non-compulsory workshops, mostly attended by PhD students (in history, the humanities, and social sciences), each participant presents their sources and questions, and we (and the group) discuss how datasets could be built and quantitative analysis could be useful. Not incidentally, questions about the adequation between sources and questions also arise quite often. Not incidentally either, the audience in our workshops has changed over the years, from mostly

16 Levi, "On Microhistory", 106.
17 Milo, "Toward an Experimental History."
18 Ruggles and Magnuson, "The History of Quantification," notice that when American historians stopped submitting quantitative papers based on historical data, Europeans and academics in other disciplines replaced them.
19 Among other places to explore these tools, beyond chapters 4-7 of our book, see for example the journal *Poetics* generally, and Franzosi and Mohr, "New Directions in Formalization."
20 Abbott, *Methods of Discovery*, Lemercier, "Abbott et la micro-histoire."
21 Abbott, *Time Matters*, Claire Zalc, "Modéliser la persecution."
22 Works cited in our book, for example, are listed here: https://www.zotero.org/clairelemercier/items/collectionKey/Y6DGTTKB We are aware of the limits of our knowledge, esp. as regards research published neither in English nor in French, and hope to be able to expand this list.
23 For details on the formats of our courses, see Lemercier, "Teaching with our Book."

male to mostly female. The topics and areas of research are extremely diverse, but a majority have sources organized around persons, i.e. they want advice on prosopography, with significant minorities interested in text analysis. We have thus come to believe that most historians could apply quantification (and would like to do it, if properly taught) to populations or samples of a few dozen to a few thousand individuals (often persons or texts, sometimes events such as lawsuits) described with a large number of variables (generally changing over time). On occasion, we have also used some of the teaching strategies that we have devised for beginning graduate students in history with students in quantitative sociology—having them experiment with non-standard ways to input and categorize data from narrative sources, images, etc. We believe that this format could also work in economics.

In France like elsewhere, most students come to the humanities and history on a "literary" track, convinced that mathematics were not for them. And as historian Antoine Prost pointed out, the social stigma "not good in math" can be turned around: "certain self-styled princes of the intellect commonly express haughty disdain for insistence on rigor or quantitative discipline of any sort, as though these were trivial concerns, menial chores to be left to subordinates."[24] Our students thus lack a role model, between the "princes of the intellect" and those social scientists (e.g. some economists) who use quantification as distinction. We try to provide them with an alternative.

Our courses and our books are different from the few examples that we knew as students, not only because they address many different methods, but mostly because we emphasize (in the courses even more than in the books) what comes before "quantitative analysis" *per se*: how to go from an archive (or any other document) to the "clean" datasets of social scientists (a .csv file, a "tidy" file in R, etc.). What we had not anticipated in our first years of teaching, at a time when we focused on introducing regression, network analysis, etc., was that those first, preliminary steps were at least as exotic for students (be they historians, sociologists, or economists) as statistics. Many had never heard about variables, never seen a .csv file; but those who knew what to do with "clean" rows and columns did not know better how to get there *from a source*. Students would either write down narrative details in their database and only use queries for analysis or, trying to emulate an implicit model of "clean data," they would directly translate their sources into supposedly unambiguous, rigid categories: 1 for upper-class, 2 for middle-class, 3 for workers, and so on. We knew the pitfalls of these two solutions, but it took us a long time to formulate alternative principles. We came to appreciate that data preparation is already an important stage of analysis, and certainly not the chore often called "cleaning" and therefore delegated to underlings.

On the basis of this experience, the remainder of this text fleshes out *which* quantitative history we want to teach; but *how* should it be taught? In this respect, we want to emphasize three ideas.[25]

First, the best way to teach and learn "quanti" is to do research together. One of the favorite metaphors in our teaching is that of assembling an Ikea piece of furniture. You can always learn the instructions by heart, but the important things happen when you are in front of the parts to be assembled; and each piece of furniture has a specific manual, even if there are some generic parts. Hence our "hands-on" approach, always starting with actual research questions and sources. In graduate workshops, each participant comes with her questions and sources and

24 Prost, *Douze leçons*,

[25] For further comparisons with teaching strategies in history and sociology, see Mercklé & Zalc, "Teaching 'Quanti'."

shares them in turn with the group, who exchange ideas as to the constitution, categorization, and analysis of data. In master courses, we give the students one source (e.g. the official biographies of French MPs in 1946-58) and they collectively decide on how to sample or focus on specific populations, to input data, to categorize it, to formulate reasonable research questions, to produce provisional answers based on correlations (contingency tables and chi-squared tests) and close reading of specific cases, and to discuss the results. We have experimented this format with historians and sociologists, but it would arguably be as difficult (and as attractive) for economists.

Second, "hands-on" involves actual historical sources (which digitization has made easier), not forged examples and simulated data (as is often the case in introductions to quantitative sociology) or aggregate, "clean" data from published syntheses. Even if the course does not allow to perform each step, from the source to the analysis and discussion, with the class, it is important to mention all steps and show what the successive files look like. And spending more time on the initial steps makes sense: if one has to resort to learning alone, following tutorials (the Ikea manual) is doable if the aim is to learn regression or network analysis on structured, categorized data. It is almost impossible, on the contrary, to learn alone how to constitute such data.

Third, it is possible to understand quantification, learn skills and develop an appetite for quantitative methods without being good in mathematics. We have to address what sociologists call "math anxiety,"[26] not reinforce it. What we teach is not statistics, but some principles derived from statistics (e.g. about sampling)—principles that can be explained in natural language– and the proper use of software based on statistics for the exploration of historical sources. We do not want to hijack history classes to train students to the mathematics that they have been disgusted with before. We want them to become informed users of quantitative methods—to understand what a method does and does not do, its advantages and pitfalls, from the extant literature; not to demonstrate the theorem behind the software. Similarly, we have no religion about the "historian-as-programmer," a creature of the 1960s recently revived by "digital historians." Sometimes, it is necessary to write a short script (in fact, to copy, paste, and edit one) because there is no simpler way to perform an analysis that is useful for a dissertation; any historian can learn how to do this. Often, however, a contingency table and chi-squared test will do.[27]

2. Data from Close Reading of Historical Sources

Almost all textbooks on quantification begin with data arranged in neat rows and columns and standardized in simple codes: there are just two genders, no unknowns, a finite number of occupations, one per individual, et cetera. It is then tempting either to consider that complicated sources are not amenable to this format, and renounce quantification, or to naïvely simplify the source in the interest of "cleanness." The lack of interest in the constitution of data leads to, then is further increased by its subcontracting to others—research assistants or even computers, left to deal with the variety found in the sources. Discussions on the topic tend to be confined to promoters of re-usable datasets who often advocate "one-size-fits-all" data structures or categorizations. On the contrary, whereas we offer general advice in our book, it includes the idea that choices made in the constitution and categorization of data should be tailored to the specific source and questions. Making these choices explicit might allow others, in the future,

26 Paxton, "Dollars and Sense", Decesare, "'Statistics Anxiety'."

[27] See the "How To" category in our blog (https://quanthum.hypotheses.org/category/how-to) for specific examples of this stance toward tools and programming.

to use our data as new sources; but these (unlikely) future uses should not constrain our choices. We therefore advise our students and readers about the critical, close reading of numerical as well as textual sources—a crucial first step of analysis—and we urge them not to confine "distant reading" to its usual targets, and especially not to the already digitized sources.

2.1 Any Source Can Be Quantified

When we teach on the basis of collective research, we begin with the source. As prosopography is a widespread practice in history and as we needed a source that was easily available online, we have often used second-hand sources: biographies compiled by historians or by institutions for memorial purposes. For example, several groups worked on the official biographies of French MPs in the 1950s.[28] In this context, we ask students to read a few biographies and think about "what type of information we could extract from these biographies and try to quantify." They generally first come up with the "usual suspects" of sociologists: gender, age, occupation. And they do not think critically about indications of these variables in the source: as the aim is to quantify, they tend to treat all information as objective. When we urge them to think of other possibilities, however, they end up discussing what could be made of, for example, the phrasing, or lack of mention, of marital status or religious affiliation, or the headgear on pictures. One of the most brilliant essays discussed the photographs of representatives of the colonial Empire in the French Parliament. The students had created several variables describing what they saw (was there a hat? a tie? etc.) and had researched the production of the pictures. This helped them to interpret correlations between the contents of the texts and the photographs—for example, to think about the intentions of those deputies who wore supposedly traditional attire on their official photographs. Other students quantified the phrasing of biographies, correlating it with their authors or dates of writing as well as with the gender or action or their subjects. We wanted them to learn that many different aspects of a source could be quantified—not just those generally thought of as quantifiable.

Likewise, we are convinced that all kinds of sources can be treated quantitatively: statistical sources, individual records, all types of texts as well as images. We thus fight the still frequent description of the usual sources of the first wave (marriage records, census forms, probate inventories, etc.) as "quantitative sources." Those are mostly textual and can be subjected to all sorts of readings. Any type of content in any type of source (or its absence) can be quantified.

What we therefore teach is that it is not its type or format that makes a source suitable for quantification (it deals with economic matters, contains numbers, or lacks an aesthetic dimension). Sources of any type that come in large quantities (or, more precisely, can be conceived as a long series of units: records, images, paragraphs, etc.) lend themselves to quantification, if only because it is difficult to keep a general view of their contents without creating a database. But the main question is: will the constitution of data from this source help answer our questions and produce fresh knowledge? It is the research question that should drive choices as to how to quantify, not the source; but without a good understanding of the source, quantification will fail.

[28] With other groups, we used the official biographies of members of the French Academies of Science, a French-language dictionary of the workers' movements, *Maitron*, and narratives about the Righteous among the Nations presented on the official Yad Vashem website—sources that have equivalents in many countries and languages.

Historical statistics are a case in point—we had many questions in our workshops, over the years, as to how to use them, and the answer matters as much, we think for "new historians of capitalisms" as for economic historians. After having been a favorite source of the first wave, too often taken at face value, statistics of the past have mostly become, in history departments, the source of a vibrant, but non-quantitative history of statistics, too little known by non-specialists.[29] As for any other source, however, it is possible to create new numbers, substantively meaningful for our own questions, from documents (here, numbers) of the past.[30] This requires external and internal source criticism, with general questions such as: Who commissioned the statistics? Who was supposed to respond to queries, and who actually did? How did the promoters of the project define their categories? Are we sure that the actual producers of answers and numbers understood them in the same way? Which quantitative techniques were used to produce aggregate numbers, what do they emphasize and what do they hide? Who subsequently used the numbers and categories and for what purpose? If we keep these questions in mind—even if we do not have definite answers to some –, it is possible to constitute new data from the old ones and to answer some of our own questions. This reconstructionist approach of statistics has especially been used by feminist scholars to overcome biases in records of female labor. The idea of "cunning historians"[31] able to "read their sources against the grain" to produce not just new narratives but also new numbers is interesting for us because many historians think of quantification as violence done to their sources by undue simplification and abstraction. Here, the historian motto of "complicating" the story, i.e. not naïvely reading the source and acknowledging subtle power dynamics, is compatible with the building of new abstractions and measurements. In this spirit, one of us used the very statistics that had been taken for accurate and objective by some historians, then studied by Joan Scott for its gender biases, to elicit new knowledge about female apprentices in nineteenth-century Paris.[32]

Fresh knowledge can also be found beyond the proverbial streetlight[33] (extant datasets, sources routinely quantified to answer our question or similar ones). This is a moving target: for example, using probate inventories to investigate consumption was once a pioneering idea; they are still useful, but their biases and limitations have become more apparent and researchers have looked for complementary sources. For us, the idea is not to complain about the fact that a source has biases (they all have) or to find "the best" source, but to assess *which questions it* could help answer.[34] The history of working hours, productivity, and female labor in early modern Europe offers a comforting example of debates that have partly evolved thanks to the invention of new sources and discussions of their biases. For example, the reading against the grain of testimonies on crimes gave indications on working hours; the systematic gathering of verbs in diverse texts expands the range of documented occupations and tasks.[35] Passionate debates on the productivity of English spinners have led to this rare, exciting occurrence: a graph comparing measurements *across types of sources* in an economic history journal—

29 A classic chapter in this spirit is Scott, "Statistical Representations of Work."
30 The trajectory of the history of accounting is similar: Rosenthal, "Quantitative Middle Ground." For an example of reconstruction: Gervais, "Mercantile Credit."
31 Sarti et al., "Introduction," 31-37.
32 Crowston et al., " Parisian Apprenticeships."
33 Rosenthal, "Seeking a Quantitative Middle Ground", 674.
34 For an excellent example in the history of consumption, see Bedell, "Archaeology and Probate Inventories."
35 Voth, *Time and Work in England,* Ågren, *Making a Living*.

accompanied by text discussing categorizations, e.g. were workhouse spinners part of "workers"?[36]

This exercise in looking for new sources and assessing what they document and what they hide is something that could be shared by economists and historians—if only each discipline rewarded it better. It is striking, for example, that economists who took pains to delineate different types of advertising in early twentieth-century newspapers, and to produce numbers by measuring column space with a ruler, had to edit the part of their working papers where they carefully described this process from the published version.[37] Non-standard sources, categorizations, and methods tend to take time, space, and increase risks of rejection in many peer-reviewed journals.

2.2 Constituting Data, not "Cleaning" it

Some economists do recognize the best historians' distinct skills in non-naïvely reading sources. According to Naomi Lamoreaux, they know "how to read texts over and over in the context of related documents […], and how to derive meaning from what was not said as well as from what was said."[38] For us, this reading is not a mere complement to the interpretation of numbers; in our view, it is rather a pre-requisite to the constitution of data. In the same spirit, Philip Hoffman, Gilles Postel-Vinay, and Jean-Laurent Rosenthal recently wrote about working on a type of credit that was recorded in no past statistics and only appeared in scattered, non-digitized archives—but proved important, even in the perspective of economists. They concluded that "Ignoring any of those elements can cause enormous problems", in reference to: "how the original historical evidence was generated, how it was preserved, and how they [the researchers] go about collecting it."[39] Economists Howard Bodenhorn, Timothy Guinnane and Thomas Mroz, discussing historical heights, likewise took pains to explain that some problems could not be solved by econometric techniques, but required researchers to "ask hard questions about potential source bias" by investigating the historical production of their sources.[40]

Accordingly, we are very cautious about the idea of "big data" in history, which could only arise from the aggregation of small datasets (even that of Hoffman et al. is not "big" in this sense). The new fashion for bigness is often based on the old, naïve idea that the many biases will cancel one another. Archaeology and ancient and medieval history, where the compilation of sources, then the digitization of compilations is allegedly more advanced than in early modern and modern history, offer many cautionary tales in this respect. For example, Søren Michael Sindbæk, a pioneer in the careful network analysis of archaeological and narrative medieval sources, warned that "'big data' is rarely good" to conclude an experiment on a large, heterogenous repository of data on maritime networks. A network visualization of this data mostly revealed patterns in archaeological knowledge—a useful result per se, but not to be confused with patterns in medieval transportation.[41]

In order to avoid such pitfalls, external source criticisms (trying to find information as to how the source was produced) is indeed useful; in addition, the first steps of quantification offer

36 Humphries & Schneider, "Losing the Thread."
37 Bignon & Flandreau, "The Economics of Badmouthing."
38 Lamoreaux, "The Future of Economic History," 1254.
39 Hoffman et al., *Dark Matter Credit,* 219.
40 Bodenhorn et al., "Diagnosing Sample-Selection Bias," 1157
41 Sindbæk, "Northern Emporia."

potential for internal source criticism. It is one of the reasons why we are firm advocates of "dirty" datasets. In our view, the fact that historical sources are complicated, heterogeneous, and scattered is what makes them interesting, and a basis for the construction of rich (deep and dense, rather than "big") data. Data constitution and categorization should take account of "outliers and weirdness."[42]

In our teaching and books, we emphasize the importance of the input phase of research—the moment when we iteratively read and copy the source in a table and make decisions on its rows and columns, without imposing categorizations. We do so not because we wish to impose an arduous and unpleasant "rite of passage" on the researcher but because we believe that this work is a key moment in any research project. Doing the input ourselves is our chance to discover the kinds of unusual things that lead to innovative results. As tedious as inputting data may be, it offers an opportunity to become really familiar with the source. Large research projects have made the delegation of input more common, for the more privileged among us, in recent years (as it was during the first wave). We consider this a risk, yet we have tried to adapt our principles to this context, because it offers the opportunity to gather original data, especially when the sources are multilingual. For example, Claire Zalc leads an ERC-funded research on the social and geographical trajectories of Jews in Lubartów, a village in Poland, in the interwar period. One of the sources is a population register from 1932, which lists all the households, apartment by apartment.[43] The constitution of data is teamwork, with a lot of discussions, and Claire Zalc is fully involved. Beyond the obvious, e.g. dates of birth, the team recorded, for example, writing tools used in the register (pencil, red ink, etc.). Descriptive statistics on this variable allowed them to understand that the inscription "expelled", in pencil in some rows, was most probably added by the Germans during the Second World War to denote the first round up of Jews in Lubartów in April 1942 and their deportation to the Bełżec extermination camp. The team also carefully copied the exact words from the source, which brought to light variations in patronymic spellings strongly correlated with religion.

The complications of the source are therefore not an obstacle to the constitution and quantitative analysis of data. On the contrary, a quantitative analysis that does not preemptively erase those complications often allows to make sense of them. Similarly, we always advise students to carefully record missing data, rather than be ashamed and hide it. Patterns in missing data may reveal aspects of the constitution of the source. "Missing" might even become substantive information in its own right when it reveals the ability to hide from the authorities or a lack of prestige in some circles. Similarly, what is often described as "errors" in the sources, as when dates of birth change or double-entry accounts are not balanced, might become an important proxy of individual strategies or abilities.[44]

Likewise, in the context of occasional teaching to quantitative sociologists, the simple instruction to "create a mock spreadsheet with at least three rows (cases) and at least six columns (variables)" from narrative or visual documents elicited fascinating discussions as to what could be done with, for example, a variable listing phone numbers given on posters. The number itself might indicate something (e.g. a geographical location), but the *presence or absence* of a phone number could also be interesting in itself—in terms of intended audience, for example. The data is not "missing", here: the fact that we don't know indicates something

[42] Hitchcock, "Big Data."
[43] For a short presentation of the source, see Zakrzewski, "The 1932 Population Register."
[44] On missing data, Mercklé & Zalc, "Can we Model Persecution?"; on contradictions in dates of birth, Boderhorn et al., "Diagnosing Sample-Selection Bias," 1164-5; on unbalanced accounts, Rosenthal, "From Memory to Mastery," 744-5.

interesting. In our courses for beginning undergraduates, such discussions on the elaboration of collective inputting instructions take a lot of time, but students enjoy them and we think of this time as perhaps the most productive in the course. The instructions must be clear, so that the inputting is consistent, but they aim at keeping the complications of the source as visible as possible for the next steps of research.[45] Even in the apparently simple situation of biographies, we encountered limit cases. Should a member of the Parliament who died just after his election and never took part in debates get his own row? (it depends if you study elections or parliamentary work!) What should we put in the columns labeled "political action?" What counts as "political"? What counts as an "action"? Even without standardizing the contents of the columns (at this stage, the students mostly input quotations from the source), a clarification of categories, based on research questions but compatible with the actual contents of the source, is in order. In our teaching, we want to make the point that quantification does not imply that we skip such "complicated" issues. On the contrary, trying to be systematic can help us to think about our often implicit definitions. This process of devising instructions based on a close reading of the text is not only useful for source criticism: it helps to clarify the very questions that the research is supposed to answer.

3. Constructivist, Experimental Quantification

This view on input is intimately linked to our practice of the further stages of quantitative history. We want input to keep the words of the source, its lacunae and inconsistencies, because we have no unique, pre-determined categorization scheme: experiments in categorization arise from the encounter between our questions and the source. The research practice that we teach thus often results in spreadsheets with hundreds of variables (including those directly copied from the source and their categorized variants, with separate columns for each source and often each date or period), whatever the number of individuals. This is only manageable because we know that different methods will allow us to explore different corners of this complicated whole: they produce several "distant readings" as well as identify interesting outliers.

3.1 Categorization: Imaginative and True to the Source

Categorization is one of the points on which critiques of quantitative methods have focused since the 1980s. The potential pitfalls are many, according to historians: anachronistic use of nomenclature, reification of individuals, and improper aggregation of diverse entities. Economists are also concerned with some of these problems, even if they rather call them bad proxies or heroic assumptions. No categorization is perfect in itself, but some are more or less well adapted to certain research objectives, whether in practical terms (such as the number of classes), theoretical terms (criteria of classification), or rhetorical terms (names assigned to the groups in the chosen scheme of classification). As statistician Alain Desrosières put it, "The question is not: 'Are these objects really equivalent?' but: 'Who decides to treat them as equivalent and to what end?'"[46] Distinguishing the input from the categorization phase is a first step if we want make the latter acceptable by the average historian, and to make it more meaningful for all, including economists.

How should I categorize occupations? This is a classical question in our workshops, as in most circles interested in quantitative history, and always a daunting one. As always, we mostly

[45] Extracts are presented in Lemercier, "Teaching with our Book."
[46] Desrosières, "How to Make Things which Hold Together," 201.

answer "it depends on your questions and on what is in your source" and elaborate on possibilities, emphasizing the fact that is will probably be useful to have more than one scheme and giving examples. Studying the Business register of Paris in the interwar period for her PhD dissertation (registration was compulsory for all entrepreneurs), Claire Zalc wanted to categorize the "business purpose" field in the form. Businesses were as diverse as souvenir stands, insurance brokerages, laundries, belt manufacturers, and fruit and vegetable wholesalers, but the precision of declarations varied widely. Some terms, such as "textiles," do not indicate whether the business is manufacturing or retailing. Claire Zalc built categories by iteratively aggregating mentions that she deemed close enough—beginning with many small classes, then merging them. Yet she kept one specific occupation apart: hairdressers. This choice was dictated by her general knowledge of the context rather than by the size of the group. The French National Association of Hairdressers indeed alleged in the 1930s that recent immigrants were particularly likely to enter the trade: it was interesting to compare this complaint with the information in the Business register. [47] In her own dissertation, Claire Lemercier focused on the elite of Parisian business in the nineteenth century. For each member of the Chamber of Commerce, separate sources for several years gave her different occupations: many were described as bankers as well as more or less specialized manufacturers and wholesale merchants. She however produced a very simple table showing changes in the shares of "bankers," "merchants," and "others" over time; the important thing here is the legend explaining, for example, that a "banker, textile manufacturer, and wine merchant" counts as a banker—because she considered that it was the best-suited definition for the question that she was asking in these pages. In a different chapter, discussing a different issue, she used a definition of *la haute banque* (merchant banking) was based on family ties rather than stated occupations.[48]

In these cases, we began with one field in the original data from one source, or several similar fields (denoting a type of business) from several sources, and we ended up with several categorization schemes. Often, definitions are based on more composite criteria. Digital humanists Miriam Posner and Marika Cifor have reported on an interesting teaching experience.[49] The aim was the creation of a database on early African-American silent "race films," so students had to decide on a definition of the genre. This could not be a preliminary decision: possibilities would depend on the available data. They settled on a definition of the "race film" as "a film with African American cast members, produced by an independent production company, and discussed or advertised as a race film in the African American press," for reasons that they take care to explain. They also kept a separate file for "discarded data" so that other scholars could make different decisions. Similarly, our own students have to decide in small groups on categorization schemes, then make them explicit for other groups. We take pains to explain that there is no intrinsically bad categorization scheme, only insufficiently explicit ones, and many that are not adapted to the research question. For example, as regards occupations in biographies, we encountered students interested in the fact that the individual had the same occupation as one of his or her parents (they had to explain what "the same" meant in limit cases), or not, or it was impossible to know from the biography.

In our book, we warn against the pitfalls of a frequent categorization scheme: the use of names as proxies of origin, ethnicity or religion.[50] We do not, however, address proxies more generally. We insist on the fact that closely reading the source during the input phase, as well

[47] Zalc, *Melting Shops*.
[48] Lemercier, *Un si discret pouvoir*.
[49] Posner & Cifor, "Generative Tensions."
[50] Lemercier & Zalc, *Quantitative Methods*, 66-67.

as using descriptive statistics or clustering techniques on not-too-simplified data, can produce interesting ideas for categorization. But we do not believe that the analytic categories could or should directly come from the historical materials: we just try to balance the interaction between preliminary questions and surprises from the source. Historians have often criticized cliometricians for their tendency to look for pre-defined entities such as GDP, "human capital" or "skilled occupations" in whatever source was most easily available, only producing numbers thanks to "heroic assumptions."[51] The economic history of Africa, when it involves measurements of population, economic activity, ethnic fragmentation, etc. in the pre-colonial period, is particularly susceptible to criticisms aimed at bad proxies.[52] This tendency to use questionable proxies has arguably been compounded by the availability of easily downloadable data. The problem is not that the assumptions are heroic but that they are, from the point of view of historians, not very explicitly stated or justified, and pave the way for over-interpretation. Of course we touch here to fundamental (and aesthetic) differences between disciplines. Whereas some historians start with a source but no explicit research question, many economists start with a well-defined question and a standard method, then go look for the most convenient source[53]. But as historians, we do welcome readings of the sources against their grain and bold experiments in the creation of proxies—as long as the rules of the interpretive game are kept explicit.[54]

3.2 Analysis: There is a Life Outside Regression

As advocates of close reading as a first stage of quantification, rich rather than big data, and the heuristics of outliers, do we therefore abdicate any attempt at generalization, or answering big questions? Absolutely not. We simply do not believe that quantitative history should or can only focus on the typical, or the average.[55] Datasets that do not involve many individuals can provide answers to big questions; shallow data on a large number of cases, produced without source criticism, should not.[56] Our questions, however, are often not similar to those of standard cliometrics paper. They are sometimes mostly descriptive—because when we do not have a good description, causal questions seem rather pointless. Visualization tools, for example, are particularly adept for tackling descriptive questions[57]. Our definition of causal questions is also more inclusive than that of standard econometrics. Hence our teaching of a wide range of tools, beyond regression.

---

[51] For a sober assessment of this problem in classical works on capitalism and slavery, see O'Sullivan, "The Intelligent Woman's Guide", 764-774, Hilt, "Revisiting *Time on the Cross*."

[52] Bourgeois-Gironde & Monnet, "Natural Experiments," Hopkins, "The New Economic History."

[53] We thank Alberto [*last name ? sorry, I did not get it - in the workshop in January*] for this observation.

[54] See, for example, how a sociologist and a mathematician built data on possible exchanges of information between captains from records of ship voyages in the East India Company: Erikson & Samila, "Networks, Institutions, and Uncertainty."

[55] As was recently stated by Hilt, "Revisiting *Time on the Cross*," after and certainly before many others.

[56] In addition to the classics of micro-history, and in fact many works in cliometrics, see for example Zalc & Bruttmann, *Microhistories of the Holocaust*, and Stephenson "'Real' Wages?" (cited, for this very reason, by Brownlow, "Economic History," 362).

[57] On the wide menu of formalization tools, see Tilly, "Observations of Social Processes." Claveau, "The Bibliometric History," uses network visualizations to renew the history of economic thought. Cunfer, "Scaling the Dust Bowl," investigated the Dust Bowl using a geographical information system and excellent source criticism. In this chapter, it is the description that changes the understanding of the event and opens new causal questions.

When we first attempted to teach quantification, in the early 2000s, we focused on "new methods." Reading sociologists, we had learned about network analysis, sequence analysis, and event history analysis. We thought that these methods were well-suited to explore what micro-historians found interesting: interacting individual trajectories. Discussions and further readings led us to appreciate that factor analysis, regression, and text analysis, however older, were also still useful for the types of data that we and our students wanted to analyze, and were therefore worth teaching. We now focus more on the constitution of data, but one thing has not changed: we always mention several methods, and we advocated the use of the one most suited to the user's data and questions, rather than the supposedly easiest, newest, or politically better one. A single quantitative method, however refined, cannot answer all the historical questions. Each one can play a role at different stages of reasoning and it is often heuristic to combine them. We fight the arms race that often prevails in digital humanities as well as the exclusive focus on regression in cliometrics: we like topic modeling and network analysis per se, but we do not think that they should be applied to all research. In our teaching, we emphasize the that mere contingency tables with chi-squared tests can answer many historical questions. But we also encourage beginners not to self-censor if their data and questions require the use of one or even several supposedly "more advanced" method. We promote knowledge of diverse tools, rather than the advanced mastery of a few—and we try to keep our eyes open for the discovery of new ones in disciplines, specialties, or countries that we do not know well. We think that all willing historians and economists, perhaps using our book as a first step, can become skilled readers of papers using many methods (even without using those in their personal research) and then teach this type of numeracy.[58]

In our own research, we have taken advantage of being able to switch between methods that offer different perspective on our data. In a paper focused on the value of this diversity of tools, Pierre Mercklé and Claire Zalc show that the successive use of multiple correspondence analysis, regression, network analysis and sequence analysis elicit different facets of the logic of persecution and survival of the Jews of Lens.[59] Doing this, they "complicate" narratives and explanations, in the positive sense of the word, in the historians' use: the narratives and explanations become more nuanced, account for more cases (not just the most frequent) and for differences between historical contexts. Do they become too complicated? Social scientists with a taste for "parsimonious" models might think so, yet this is still quantification—data has been abstracted and hypotheses made explicit. We know that some economists—but few—have added tools other than variants of regression, including "descriptive" ones, to their toolbox; for example, Marc Flandreau and Clemens Jobst have been among the first non-sociologists to use network analysis on historical data, in a paper that also made an inventive use of the press as a source and carefully discussed its proxies.[60]

Discussions around papers like those—neither standard qualitative historical papers not standard quantitative papers in economics, yet distinctly recognizable as belonging to history and to economics –, even if participants are few, could make our respective scientific and aesthetic criteria clearer. There would hopefully be common ground around the constitution and perhaps the categorization of data. The choice of methods and the status of models might

[58] For some ideas of reading lists, see the category "Good Reads" in our blog (https://quanthum.hypotheses.org/category/good-reads)

[59] Mercklé & Zalc, "Can we Model Persecution?" For a joint use of regression and multiple correspondence analysis, see also François & Lemercier, "Financialization French-Style".

[60] Flandreau & Jobst, "The Ties that Divide". A recent introduction to network analysis for economic historians shows that the field is finally blossoming (many years after it has in sociology and in other parts of history, and in relative isolation from those): Geisler Mesevage, "Network Analysis."

remain more distinctive of each discipline, but we would like, at least, to disentangle substantive from institutional reasons (the standards of journals and hiring committees) for this difference. Cliometricians have invested a lot in a very peculiar definition of causation—a definition that is not even shared by all economists, especially in macro-economics.[61] Historians benefit from a less rigid standard, in that they can more easily experiment; the drawback is that they rarely pause to wonder about what exactly they call a cause. Amicable but frank discussions with economists and other social scientists could improve this situation—which is all the more important at a time when some "data scientist" advocate for dropping causal questions altogether.

Conclusion

We would like to end this paper by addressing data, rather than quantification per se. Data is now the fashionable term, but more importantly, our teaching led us to conclude, like many before us, that good data was key to good quantitative research in history. The Social Science History Association, part of the surviving offspring of the first wave of quantitative history, recently issued a call for papers on "Data and its Discontents." We would like to offer our answer: the one that we have been teaching for years, and have had the pleasure to recognize in many papers and books by our former students. It is originally intended for the average historian, but perhaps also interesting for economists.[62]

Data exists; it is as simple as notes taken from historical sources. We tend to call it "data" when we take our notes in a spreadsheet, but it is really any type of notes taken systematically from historical sources. Systematically, but not automatically, mind you: the idea is to give oneself a set of rules suited to the source and questions, to make them explicit, and to follow them consistently—not to defer to a computer for the reading of the source. Our spreadsheets are verbose and most "digital humanists" would want to "clean" them, but we cherish their dirtiness. It is a sign that the transformation from source to data (notes) happened through our own thought processes (not those of underlings paid to sweep the dirt under the rug) and retained the ambiguities of the sources. To make data better, and to have more colleagues crave it and fewer abhor it, we want to keep it as close as possible to the source, even though it will be dirtier, and more costly to produce in large quantities. But that's fine, because we do not believe data is good only if it is really big. It is good if it is complicated, and thus rich in information, but still systematically acquired and noted in a structured way, so that we can simplify it in many different ways if we want to experiment with it. It is good for thinking, even, or especially, when it produces new questions rather than final answers.

[61] Bourgeois-Gironde & Monnet, "Natural Experiments," Hopkins, "The New Economic History." For thought-provoking discussions of causation in sociology, political science, and interdisciplinary contexts, see also Abbott, *Time Matters*, Ragin, *Redesigning Social Inquiry*, Grenier et al., *Le modèle et le récit*. On misunderstandings about the meaning of "cause" and counterfactuals between economic historians and new historians of capitalism, see Blin & Barreyre, "À la redécouverte du capitalisme."

[62] We thank Clare H. Crowston, who initially co-wrote the last paragraph with Claire Lemercier for the SSHA Conference, Chicago, 2019. For the call for papers, see https://ssha.org/files/2019%20SSHA%20CFP.pdf

References

Abbott, Andrew D. *Time Matters. On Theory and Method*. Chicago: University of Chicago Press, 2001.

Abbott, Andrew D. *Methods of Discovery: Heuristics for the Social Sciences*. New York: W.W. Norton & Co, 2004.

Ågren, Maria, ed. *Making a Living, Making a Difference. Gender and Work in Early Modern European Society*. Oxford: Oxford University Press, 2017.

Armitage, David, and Jo Guldi. *The History Manifesto*. Cambridge: Cambridge University Press, 2014. https://www.cambridge.org/core/what-we-publish/open-access/the-history-manifesto.

Bedell, John. "Archaeology and Probate Inventories in the Study of Eighteenth-Century Life." *The Journal of Interdisciplinary History* 31, no. 2 (2000): 223–45.

Bignon, Vincent, and Marc Flandreau. "The Economics of Badmouthing: Libel Law and the Underworld of the Financial Press in France Before World War I." *The Journal of Economic History* 71, no. 3 (2011): 616–53.

Blin, Alexia, and Nicolas Barreyre. "À la redécouverte du capitalisme américain." *Revue d'histoire du XIXe siècle*, no. 54 (2017): 135–48.

Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz. "Diagnosing Sample-Selection Bias in Historical Heights: A Reply to Komlos and A'Hearn." *The Journal of Economic History* 79, no. 4 (2019): 1154–75.

Boldizzoni, Francesco. *The Poverty of Clio. Resurrecting Economic History*. Princeton, N.J.: Princeton University Press, 2010.

Bourgeois-Gironde, Sacha, and Éric Monnet. "Natural Experiments and Causality in Economic History: What Relation to Theory and Temporality?" *Annales HSS* 72, no. 4 (2017): English version forthcoming.

Brownlow, Graham. "Economic History, the History of Economic Thought and Economic Policy." In *An Economist's Guide to Economic History*, edited by Matthias Blum and Christopher L. Colvin, 47–53. Palgrave Studies in Economic History. Cham: Springer International Publishing, 2018.

Cellier, Jacques, and Martine Cocaud. *Traiter des données historiques. Méthodes statistiques, techniques informatiques*. Rennes: Presses universitaires de Rennes, 2001.

Claveau, François. "The Bibliometric History of Specialties in Economics." Digital History of Science, s.n. http://www.digitalhistoryofscience.org/economics/index.php.

Crowston, Clare H., Steven L. Kaplan, and Claire Lemercier. "Parisian Apprenticeships in the Eighteenth and Nineteenth Centuries." *Annales HSS* 73, no. 4 (2018): English version forthcoming.

Cunfer, Geoff. "Scaling the Dust Bowl." In *Placing History: How Maps, Spatial Data and GIS Are Changing Historical Scholarship*, edited by Anne Kelly Knowles and Amy Hillier, 96–121. Redlands, Ca.: ESRI Press, 2008.

Decesare, Michael. "'Statistics Anxiety' Among Sociology Majors: A First Diagnosis and Some Treatment Options." *Teaching Sociology*, no. 65 (2007): 360–67.

"Debating the *Longue Durée*." *Annales. Histoire, Sciences Sociales* 70, no. 2 (2015): 285–378.

Desrosières, Alain. "How to Make Things Which Hold Together: Social Science, Statistics and the State." In *Discourses on Society: The Shaping of the Social Science Disciplines*, edited by Peter Wagner, Björn Wittrock, and Richard P. Whitley, 195–218. Dordrecht: Kluwer Academic Publishers, 1991.

Erikson, Emily, and Sampsa Samila. "Networks, Institutions, and Uncertainty: Information Exchange in Early-Modern Markets." *The Journal of Economic History* 78, no. 4 (December 2018): 1034–67.

Flandreau, Marc, and Clemens Jobst. "The Ties That Divide: A Network Analysis of the International Monetary System. 1890-1910." *Journal of Economic History* 65, no. 4 (2005): 977–1007.

François, Pierre, and Claire Lemercier. "Financialization French-Style (1979-2009). Changes in Large Firms and Conversion of Elites." *Revue Française de Sociologie* 57, no. 2 (2016): 269–320.

Franzosi, Roberto, and John W. Mohr, eds. "New Directions in Formalization and Historical Analysis." *Theory and Society* 26, no. 2/3 (1997): 133–391.

Geisler Mesevage, Gabriel. "Network Analysis." In *An Economist's Guide to Economic History*, edited by Matthias Blum and Christopher L. Colvin, 433–41. Palgrave Studies in Economic History. Cham: Springer International Publishing, 2018.

Gervais, Pierre. "Mercantile Credit and Trading Rings in the Eighteenth Century." *Annales HSS* 67, no. 4 (2012): 731–63.

Grenier, Jean-Yves, Claude Grignon, and Pierre-Michel Menger, eds. *Le modèle et le récit*. Paris: Editions de la Maison des sciences de l'homme, 2001.

Hilt, Eric. "Revisiting *Time on the Cross* After 45 Years: The Slavery Debates and the New Economic History." *Capitalism*, forthcoming.

Hitchcock, Tim. "Big Data, Small Data and Meaning." *Historyonics* (blog), November 9, 2014. http://historyonics.blogspot.com/2014/11/big-data-small-data-and-meaning_9.html.

Hoffman, Philip T, Gilles Postel-Vinay, and Jean-Laurent Rosenthal. *Dark Matter Credit: The Development of Peer-to-Peer Lending and Banking in France*. Princeton: Princeton University Press, 2019.

Hopkins, A. G. "The New Economic History of Africa." *The Journal of African History* 50, no. 02 (2009): 155–77.

Karila-Cohen, Karine, Claire Lemercier, Isabelle Rosé, and Claire Zalc, eds. "Quantitative History." *Annales HSS* 73, no. 4 (2018): English version forthcoming.

Lamoreaux, Naomi. "The Future of Economic History Must Be Interdisciplinary." *The Journal of Economic History* 75, no. 4 (2015): 1251–57.

Lemercier, Claire. *Un si discret pouvoir. Aux origines de la Chambre de commerce de Paris, 1803-1853*. Paris: La Découverte, 2003.

Lemercier, Claire. "Abbott et la micro-histoire : lecture croisée." In *Andrew Abbott et l'héritage de l'école de Chicago*, edited by Didier Demazière and Morgan Jouvenet, Vol. 2: online. Editions de l'EHESS, 2016.

Lemercier, Claire. "How Should I Read This Book?". *Quantitative Methods in the Humanities* (blog), February 28, 2019. https://quanthum.hypotheses.org/62.

Lemercier, Claire. "Teaching with Our Book." *Quantitative Methods in the Humanities* (blog), August 9, 2019. https://quanthum.hypotheses.org/254.

Lemercier, Claire, and Claire Zalc. *Méthodes quantitatives pour l'historien*. Paris: La Découverte, 2008.

Lemercier, Claire, and Claire Zalc. *Quantitative Methods in the Humanities. An Introduction*. Translated by Arthur Goldhammer. Charlottesville: University of Virginia Press, 2019.

Levi, Giovanni. "On Microhistory." In *New Perspectives on Historical Writing*, edited by Peter Burke, 93–113. Cambridge: Polity Press, 1991.

Mercklé, Pierre, and Claire Zalc. "Teaching 'Quanti.' Lessons from French Experiences in Sociology and History." *Bulletin de Méthodologie Sociologique / Bulletin of Sociological Methodology*, no. 136 (2017): 40–52.

Mercklé, Pierre, and Claire Zalc. "Can We Model Persecution? The Contributions and Limitations of Quantitative Methods Applied to the Shoah." *Annales HSS* 73, no. 4 (2018): English version forthcoming.

Milo, Daniel S. "Toward an Experimental History, or Gay Science." *Strategies*, no. 4–5 (1991): 87–126.
O'Sullivan, Mary. "The Intelligent Woman's Guide to Capitalism." *Enterprise & Society* 19, no. 4 (2018): 751–802.
Paxton, Pamela. "Dollars and Sense: Convincing Students That They Can Learn and Want to Learn Statistics." *Teaching Sociology*, no. 34 (2006): 65–70.
Posner, Miriam, and Marika Cifor. "Generative Tensions: Building a Digital Project on Early African American Race Film." *American Quarterly* 70, no. 3 (2018): 709–14.
Prost, Antoine. *Douze leçons sur l'histoire*. Paris: Éditions du Seuil, 1996.
Putnam, Lara. "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast." *The American Historical Review* 121, no. 2 (2016): 377–402.
Ragin, Charles C. *Redesigning Social Inquiry. Fuzzy Sets and Beyond*. Chicago: Chicago University Press, 2008.
Rosental, Paul-André. "Construire le 'macro' par le 'micro' : Fredrik Barth et la microstoria." In *Jeux d'échelle. La micro-analyse à l'expérience*, edited by Jacques Revel, 141–61. Paris: Gallimard, 1996.
Rosenthal, Caitlin C. "From Memory to Mastery: Accounting for Control in America, 1750–1880." *Enterprise and Society* 14, no. 4 (2013): 732–48.
Rosenthal, Caitlin. "Seeking a Quantitative Middle Ground: Reflections on Methods and Opportunities in Economic History." *Journal of the Early Republic* 36, no. 4 (2016): 659–80.
Ruggles, Steven, and Diana L. Magnuson. "The History of Quantification in History: The JIH as a Case Study." *The Journal of Interdisciplinary History* 50, no. 3 (2019): 363–81.
Sarti, Raffaella, Anna Bellavitis, and Manuela Martini. "Introduction." In *What Is Work? Gender at the Crossroads of Home, Family, and Business from the Early Modern Era to the Present*, edited by Raffaella Sarti, Anna Bellavitis, and Manuela Martini, 1–84. New York: Berghahn Books, 2018.
Sewell, William H. "The Political Unconscious of Social and Cultural History, or, Confessions of a Former Quantitative Historian." In *Logics of History: Social Theory and Social Transformation*, 22–80. Chicago: University of Chicago Press, 2005.
Sindbæk, Søren Michael. "Northern Emporia and Maritime Networks. Modelling Past Communication Using Archaeological Network Analysis." In *Harbours and Maritime Networks as Complex Adaptive Systems*, edited by Johannes Preiser-Kapeller and Falko Daim, 105–17. Mainz: Verlag des Römisch-Germanischen Zentralmuseums, 2015.
Stephenson, Judy Z. "'Real' Wages? Contractors, Workers, and Pay in London Building Trades, 1650–1800." *The Economic History Review* 71, no. 1 (2018): 106–32.
Stone, Lawrence. "The Revival of Narrative: Reflections on a New Old History." *Past & Present*, no. 85 (1979): 3–24.
Scott, Joan W. "Statistical Representations of Work: The Chamber of Commerce's Statistique de l'Industrie à Paris, 1847-48." In *Work in France : Representations, Meaning, Organization, and Practice*, edited by Steven L. Kaplan and Cynthia J. Koepp, 335–63. Ithaca: Cornell University Press, 1986.
Tilly, Charles. "Observations of Social Processes and Their Formal Representations." *Sociological Theory* 22, no. 4 (2004): 595–602.
Trivellato, Francesca. "Is There a Future for Italian Microhistory in the Age of Global History?" *California Italian Studies* 2, no. 1 (2011): online.
Trivellato, Francesca. "Microstoria/Microhistoire/Microhistory." *French Politics, Culture & Society* 33, no. 1 (2015): 122–34.
Voth, Hans Joachim. *Time and Work in England, 1750-1830*. Oxford: Clarendon Press, 2000.

Weingart, Scott B. "Punched-Card Humanities. Lessons From Digital History's Antecedents." *The Scottbot Irregular* (blog), October 30, 2016. https://scottbot.net/lessons-from-digital-historys-antecedents/.

Zakrzewski, Franciszek. "The 1932 Population Register." *Lubartworld* (blog), 2020. https://lubartworld.cnrs.fr/en/register-1932/.

Zalc, Claire. "Immigrants et indépendants. Parcours et contraintes. Les petits entrepreneurs étrangers du département de la Seine (1919-1939)." PhD dissertation in history, Université de Paris X-Nanterre, 2002.

Zalc, Claire. *Melting shops. Une histoire des commerçants étrangers en France*. Paris: Perrin, 2010.

Zalc, Claire. "Modéliser la persécution. Les apports d'Abbott à l'écriture historienne." In *Andrew Abbott et l'héritage de l'école de Chicago*, edited by Didier Demazière and Morgan Jouvenet, 1:419–40. Editions de l'EHESS, 2016.

Zalc, Claire, and Tal Bruttmann, eds. *Microhistories of the Holocaust*. New York: Berghahn Books, 2016.