Introduction: theory and econometrics in historical analysis

• Gabriele Cappelli
  University of Siena
  https://orcid.org/0000-0001-8934-953X
  gabriele.cappelli@unisi.it

• Nikos Benos
  University of Ioannina
  https://orcid.org/0000-0002-9361-8808
  nbenos@uoi.gr

• Yorgos Goletsis
  University of Ioannina
  https://orcid.org/0000-0002-6227-0810
  goletsis@uoi.gr

Abstract

This article introduces the special issue of the Revista de Historia Industrial – Industrial History Review on theory and econometrics in historical analysis. First, it briefly reviews the literature on the evolution of economic history since the so-called cliometric revolution, i.e., the growing use of theory to frame historical evidence and the increasing use of quantitative techniques, particularly to study causality. Secondly, it provides an overview of how a potential data revolution – prompted by technological progress in historical-data collection and processing – may be affecting the field. Lastly, the introduction provides a succinct summary of the articles included in the special issue, which highlight how economic theory and new quantitative analyses can be used to explore a range of issues within economic history, spanning different countries and time periods.

Keywords: cliometrics, cliodynamics, economic history, theory, historical analysis, data revolution.

JEL Codes: A12, B25, B41, N01.

1. This introduction is to be considered as a guest editors’ note and, therefore, has not been submitted to external review. Nevertheless, it has been reviewed by the Editorial Board of RHI-IHR.
1. Introduction

This special issue of the Revista de Historia Industrial – Industrial History Review includes articles that rely on theory and econometric analysis to improve our understanding of the past. The so-called cliometric revolution, which started in the US more than half a century ago (Haupert 2019; Diebolt and Haupert 2019), has now spread to different countries and affected the way research in economic history is conducted. The term cliometrics (stemming from Clio, the muse of history) indicates the growing use of theory to frame and discuss historical sources and evidence pertaining to the economy in the past, as well as the increasing use of quantitative techniques to answer questions related to “knowing the causes of things”, particularly in relation to constructing plausible and valid counterfactuals in social and economic history.²

Despite its importance for the field of economic history nowadays, cliometrics has not always been part of it – while its relationship with economists and historians has evolved and changed since the 1960s. The introduction to this special issue, titled “Theory and econometrics in historical analysis”, provides a brief overview of the relationship between the cliometric approach to economic history and related fields (mostly economics and history) up to the present day. Although this brief (and possibly non-exhaustive) review does not aim to shed new light on the relationship between economic history, on the one hand, and economic theory and econometrics, on the other hand, we wish to summarise existing evidence on the matter while introducing the reader to the content of this issue. Section 2 describes the relationship between economic history and related fields – economics, history, and other disciplines. Section 3, instead, discusses how a potential data revolution will affect the field. Section 4 concludes by introducing the reader to the articles which are part of the special issue.

2. The relationship between economic history and related fields

2.1. Economic history and economics

According to Cioni, Federico, and Vasta (2021), since the early 2000s, a new revolution within the economic history field has gained momentum, namely the diffusion and success of so-called persistence studies, i.e., research

² This paradigm is well represented by the LSE’s motto, “Rerum cognoscere causas”, which highlights the importance of understanding causality mechanisms governing social, political, and economic change.
that “investigates present-day outcomes as the effects of permanent features or of specific events that happened many decades or several centuries earlier” (Cioni, Federico and Vasta 2021, p. 24; Cioni, Federico and Vasta 2022). Although the last two decades have witnessed a further diffusion of the use of theory and econometrics in economic history work, the authors are sceptical about claims that see a growing integration of the fields of economics and economic history (Margo 2018, 2021). By comparing articles on economic history topics (including persistence studies) in eight leading economics journals with the articles in the top five field journals in economic history, the authors analyse similarities and differences in authorship, articles, and citation patterns. Cioni, Federico and Vasta (2021) argue that the relationship between economics and economic history is not as good as the growing use of theory and quantitative methods within the latter would have us thinking. Indeed, while interest in economic history among economists has, indeed, grown significantly, the focus of research articles in economics vs economic history journals differs quite substantially, as does the profile of authors – a very limited number of them actually engage with both economics and economic history journals and issues, even though the integration between economics and economic history displays substantial differences between North America and Europe, being stronger in the former (Fernández-de-Pinedo, La Parra-Perez and Muñoz 2023).

Cioni, Federico and Vasta’s (2021) perspective echoes that of Abramitzky (2015). The author had previously argued that economic historians should keep on placing themselves at the junction of history and the social sciences, contributing to economics, as much as demography and political science, with their own methods and focuses. Yet, Abramitzky sees economic history as a subfield of economics, which is likely to grow in the future as new technologies and the declining cost in data collection and processing imply a greater use of historical data. Along the same lines, the book edited by Bloom and Colvin (2018) serves as a comprehensive guide to economic history for economists. It covers various topics, methodologies, and debates within the field

3. Persistence studies are articles that typically deal with present outcomes but explain them as the consequences of specific events in the distant past, i.e., at least a century before the outcome (even dating back to prehistory). Perhaps the most famous example is the highly cited article by Acemoglu, Johnson and Robinson (2001) on the colonial origins of development.

4. The eight economics journals include the so-called top 5, i.e., American Economic Review (AER), Econometrica (ECMA), the Journal of Political Economy (JPE), the Quarterly Journal of Economics (QJE), and the Review of Economic Studies (RESTUD), as well as ‘other economics journals’- Economic Journal (EJ), the Journal of Economic Literature (JEL), and the Review of Economics and Statistics (RESTAT). The five top economic history journals are Cliometrica (CLIO), Explorations in Economic History (EEH), the Economic History Review (EHR), the European Review of Economic History (EREH) and the Journal of Economic History (JEH).
Introduction: theory and econometrics in historical analysis

of economic history. The book provides a valuable resource for economists seeking to engage with historical perspectives and understand the relevance of economic history in shaping economic theories and policies. As both the Great Recession and the Covid-19 crisis have clearly shown, economists do not just engage with the past by studying persistence. Instead, economists do engage with economic history in terms of “change”, when abrupt change occurs in the present shaking the equilibria of our social and economic systems, and little evidence on how such shocks may unfold in the nearest future is available (see, e.g., Clay, Lewis and Severnini 2018; Velde 2022).

2.2. Economic History and History

The mutual connection, exchange, and alleged integration of economic history and economics has been widely debated but must be somewhat acknowledged. Nevertheless, historians have been less enthusiastic than economic historians and economists about the evolution of economic history since the cliometric revolution. More than a decade ago, Boldizzoni (2011) discussed the “poverty of Clio” in his homonymous book, quite well known and debated at the time, particularly among historians. Boldizzoni’s critique of the cliometrics revolution is based on the view that neoclassical economic theory has been too predominant within economics and cliometrics, as well as the fact that cliometricians may have been quite untidy with the use of primary sources (Brownlow 2012). Although Boldizzoni confronted the cliometric revolution with a harsh tone, his call for a more diversified approach to economic history, like using theories alternative to the neoliberal or institutional paradigms or a tighter genuine connection with sociology and anthropology, are recognised to be important, albeit somewhat marginal, particularly among the practitioners of persistence studies.

Criticism of the way that the field of economic history has evolved since the cliometric revolution has come from scholars engaged with both economic history and economics, albeit more sporadically. One example is McCloskey’s (2017) discussion of the way that Fogel and North influenced the field up to the present day (McCloskey 2018; see also Diebolt and Haupert 2018). According to McCloskey, the “measurement” of the economy should be approached with careful consideration, incorporating insights from the humanities. McCloskey does not argue for a reduction in the use of theory and/or mathematics in the field of historical economics, yet she maintains that it is

5. The term historical economics is sometimes used similarly to economic history. However, while the latter is understood to indicate research aimed at understanding the past and challenging existing historiography, the former normally concerns research that looks at history to improve our understanding of the present.
crucial that the tools that economic history uses are relevant and meaningful in the context of the actual economic world. Merely proving the existence of a competitive equilibrium based on certain axioms does not contribute anything scientifically valuable – likewise, one may add, neither do proving that “history matters” or that political and economic institutions show a tendency to persist over time. What matters is the magnitude of the evidence provided that concerns certain phenomena. To put it succinctly, McCloskey argues that historical economics needs to move past the unfortunate division between facts and values, and science and humanities. Peter Temin, interviewed for the edited book *Reflections on the Cliometrics Revolution*, recognised, how, in the early days of the rise of the new economic history, the focus on economic theory and re-assessing historical narratives with new data disappointed traditional historians, and gave birth to a strong divide between practitioners of economic history in economics vs history departments (Brown 2008).

**2.3. Economic History and other disciplines**

Mejía (2015) offers a more balanced – and perhaps optimistic – perspective on the state of economic history, acknowledging its enduring influence within the field of economics over the past few decades, particularly in the last twenty years. The author contends that, in recent decades, diverse and competing agendas have been converging, forming what can be identified as an emerging or evolving paradigm that can be labelled “cliodynamics”. In contrast to cliometrics, cliodynamics offers a more flexible approach to studying history, drawing closer connections to fields like palaeontology, archaeology, and population genetics. It focuses on the mechanisms driving historical evolution rather than solely relying on data, emphasizing qualitative research and speculation to comprehend observed developmental patterns. One example of cliodynamics is provided by Grinin, Markov and Korotayev (2013), who use mathematical modelling to analyse similarities and differences between biological and social macroevolution, suggesting – to some extent – the presence within the biosphere of an analogue of the collective learning mechanism. Another example is the dataset by Turchin et al. (2012), the Historical Database of Sociocultural Evolution, which is used by the authors to investigate the origin of human ability to cooperate in large groups of individuals that are not bonded by genetics.

While cliodynamics maintains the scientific ambitions of early cliometrics by testing general theoretical descriptions of history using appropriate quantitative tools and historical data, there are noteworthy differences between the two approaches. Notably, cliodynamics emphasizes the role of ecological context, particularly the biological determinants of human interaction during periods when human groups were more susceptible to the natural environment.
Due to the challenge of measuring several of these aspects directly or easily, qualitative research and debates on the mechanisms behind observed patterns become crucial, distinguishing cliodynamics from the increasingly data-driven approach of cliometrics. Additionally, cliodynamics exhibits a broader breadth and scope compared to cliometrics. It incorporates simulations, mathematical tools, and a theoretical framework that extends beyond traditional economic theory, including game theory, institutional analysis, and recent advancements in behavioural economics. This shift in conceptual vision is also reflected in the revitalization of non-quantitative data, which aligns closely with cliodynamics in the research agenda of economic history. This resurgence has been made possible by the adoption of more flexible empirical tools, such as nonlinear models, time series analysis, and simulation methods. Indeed, even if Mejía (2015) does not explicitly address this issue, the data revolution currently affecting the social sciences (and economic history, specifically) will perhaps bring the field towards his definition of cliodynamics even further. Furthermore, as Crafts (2012) argued, economists will engage more with economic history, only insofar as economic historians can and will engage with economics and present their research results by showing the relevance and potential of their findings for policy-related questions, not just their understanding of the past.

3. A new (data) revolution?

The topic of the evolution of cliometrics, briefly reviewed and discussed in the previous sections, brings about the discussion of another revolution with the potential to reshape the field in the coming years, namely a data revolution – a potential exponential increase in the availability of historical data due to new advances in technology. The concurrent rise in the extent of digitised primary sources and data, the evolution in technologies that allow generating datasets from historical sources, the rise of AI and of the computation of qualitative sources (e.g., web and text scraping) is changing the field, by bringing about new information, and making it possible to tackle issues and questions that might not have been researchable – at least, not within a reasonable amount of time and limited resources – in the past decades. As much as the evolution of regression analysis and software has accompanied the spread of the cliometric revolution in the late twentieth century, this data revolution has the potential to reshape the field, not just in terms of methodology, but also in terms of aims and scope. If assembling large, comparative datasets concerning with historical time series and long-term development was a daunting task in the past, and has often involved a large number of scholars and universities – like the Maddison Project (Bolt and van Zanden
Gaibriele Cappelli, Nikos Benos, Yorgos Goletsis

2020), ClioInfra (OECD 2014, 2021), and the Jordà-Schularick-Taylor Mac
rohistory Database (Jordà et al. 2019; Jordà et al. 2021), to mention a few –
today’s advances in digitisation and data computing and elaboration will likely
further change the way historical data is produced and analysed.6

Extensive discussions on the implications of “big data” in the field of eco
nomic history have been undertaken by Gutmann, Merchant and Roberts
(2018). The authors delve into the meaning of this term for economic histo
rians and explore the potential opportunities that will arise in the coming
years regarding big data within the discipline. The integration of analog data,
comprising text and images, has the potential to revolutionize the field as it
can be converted into machine-readable formats. A defining characteristic
of modern big data is that the volume of stored information surpasses human
analytical capacity and tests the limits of currently available computing pow
er, thus continuously expanding the magnitude of what is considered “big”.
Today, it has become increasingly common to work with datasets comprising
40 million records or more, whether obtained from a single source or aggre
gated in a standardized format from various sources.

Economic historians have already started leveraging big data in the do
mains of population and the environment. While the collection and analysis
of big data are not novel endeavours, because governments have conducted
comprehensive censuses since the eighteenth century, compiling and publish
ing results for the entire population. However, these datasets have often been
collected and organized at the national, regional, or local levels, leading to a
reduction in the volume and diversity of information compared to the origi
nal source. Notably, economic historians have increasingly used IPUMS data,
employing the back-casting method that relies on the age structure of the pop
ulation to infer historical well-being conditions for regions such as Sub-Sahar
an Africa where systematic and representative census data is scarce until the
latter half of the twentieth century (see, e.g., Joerg Baten et al. 2021). Simul
taneously, the North Atlantic Population Project (NAPP) has provided ex
tensive individual-level data from historical censuses during the eighteenth
and nineteenth centuries (Gutmann, Merchant and Roberts 2018, p. 272). In
their extensive discussion, Gutmann, Merchant and Roberts elucidate the
benefits of employing complete individual-level records (100-percent sam
ples) to examine population dynamics and causality across various aspects of
society and the economy in the past, particularly highlighting the importance

6. Such projects are also important as they allow data to be shared with the scholarly
community and replications of previous analyses to be performed, something which has been
quite difficult until the last decade. Indeed, in the last few years, economic history journals are
increasingly adopting a policy that allows replication, requiring authors (perhaps after an em
bargo) to share their main dataset and replication files, following the “credibility revolution”
discussed by Angrist and Pischke (2010).
of capturing minority populations. While a representative 5-percent sample may suffice for studying 99 percent of the population statistically, smaller yet representative samples may not adequately capture the visibility of minority groups constituting 1 percent of the overall population.

While population data has long been considered a form of big data, other types of information are poised to assume this status in the coming years, as posited by Gutmann, Merchant and Roberts (2018). Environmental data, for instance, are becoming available with finer levels of geographic and temporal detail. These data possess strong potential to be classified as big data due to their capacity for extensive coverage, high frequency over time, and granular geographic resolution. Prominent examples include weather records spanning various parts of the world since the nineteenth century and land cover data systematically documented in the United States through aerial photographs since the 1930s and globally through satellites since the 1970s. Weather and climate data hold promise in economic history research. Analysis of high-resolution weather and climate data is already underway, shedding light on the history of natural disasters and their interplay with demographic and economic outcomes. Studies focusing on the impact of drought, for example, increasingly rely on large-scale data sources (see Rönnbäck 2014 for precolonial West Africa).

According to Gutmann, Merchant and Roberts (2018), the future of big data will encompass high-resolution images as well. Although the social sciences have made limited use of such images as a data collection method, recent studies indicate a growing trend in their adoption. Textual datasets are also gaining importance and popularity due to their ability to transform qualitative information into quantitative data. Recent research by Koschnick (2023) studies the teacher effect on students’ future research at the English universities of Oxford and Cambridge during the time of the English Scientific Revolution, by matching university students and university teachers to published works in the English Short Title Catalogue (ESTC) for the seventeenth and eighteenth centuries. Based on the subject classification entries from the ESTC, the paper uses machine learning techniques to assign topic headings to all titles within the ESTC.

Kerby, Moradi and Odendaal (2022) focus on 500 years of African economic history by analysing a corpus of texts containing more than 230,000 pages that was produced by explorers and travellers. The authors employ computational linguistic techniques such as a structural topic model approach (STM) in combination with knowledge of African economic history, to analyse how first-hand accounts (topics) evolve across space, time and traveller occupations. As the authors argue, systematically analysing such accounts through the elaboration of qualitative information through new datasets based on text will allow shedding new light on some of Africa’s precolonial past.
4. “Theory and econometrics in historical analysis”

During the last few years, cliometrics has affected the economic history of developing regions and countries, as researchers rely on a range of theories and theoretical frameworks to understand how Asia started to catch up with the West after the Great Divergence – or why Sub-Saharan Africa has not, at least with the same rapid and sustained pace. Indeed, the evolution of quantitative methods and their application to social and economic history have allowed the rapid development of new research on regions that, for a long time, have remained quite marginalised within the field, like Sub-Saharan Africa (Fourie 2016; Jedwab, Meier zu Selhausen and Moradi 2022).

This special issue has been conceived in the spirit of collecting articles that rely on theory and/or econometric evidence, though avoiding the adoption of a perspective centred on the “compression of history” (Austin 2008), i.e., so-called persistence studies (Cioni, Federico and Vasta 2021). Most of articles that are part of it were first discussed and presented at the 8th International Ioannina Meeting on Applied Economics and Finance (22-24 June 2022) in Kefalonia (Greece) during the double session on “Historical Perspectives on Industrial, Economic and Financial Development”.

Some of the articles rely on existing social and economic theory (in particular, other than neoclassical economic theory), while others use data and econometrics to investigate the issues that they are concerned with. On the one hand, new historical data – some collected ex novo from quantitative primary sources – are presented concerning the extent and change in frictions during the industrialization of Russia and the Soviet Union (1885–1940), as well as numeracy (human capital accumulation) in the regions of the former Ottoman Empire (particularly Greece) and its social and economic determinants in the nineteenth and twentieth centuries.

On the other hand, several of the articles in this special issue address the way that different theories can be used to frame and analyse the past as well as medium/long-term development. Along this line, class dynamics and distributional conflicts can contribute to explaining different phases of economic development and pace of economic growth in the economies of East Asia. Similarly, Ostrom’s design principles for the governance of common-pool resources can be used to study the importance and success of the Byzantine Monastery and its long-lasting evolution.

A brief description of the articles can be found below. We believe that the readership of the RHI-IHR will find the articles pertinent to the view that economic history and economics can be further integrated into one another in the coming future – despite recent findings that, as of today, the relationship between the two fields remains, more often than not, rather cordial and yet far from intimate.
The first article of the special issue, by Kleoniki Alexopoulou and Joerg Baten (Alexopoulou and Baten 2023), investigates how socio-economic and cultural factors shaped human capital formation across the regions of modern Greece by using untapped primary sources. The authors estimate human capital levels and trends by relying on numeracy based on the ABCC index—which measures age heaping in different populations (A’Hearn, Baten and Crayen 2009). The article discusses how education policy during the Ottoman Empire and the making of the Greek state might have shaped the aggregate trend of human capital accumulation in Greece, while econometric analysis is used to explore the relationship between numeracy and several potential determinants, particularly gender equality as well as geography and culture, i.e., the “small island effect”. Specifically, the fact that knowledge about shipping and trade requires certain quantitative skills, combined with the incentive for women to acquire numeracy as men were often away due to maritime voyages, are found to boost numeracy. According to the authors, these findings provide support for the argument by De Moor and van Zanden (2010) that “girl power” mattered for development (Carmichael et al. 2016; Baten and de Pleijt 2018).

The second article of this issue, by Guillem Blasco-Piles and Federico Tadei (Blasco-Piles and Tadei 2023), expands on previous work by Cheremukhin et al. (2017). The authors analyse the structural transformation of Russia in the period 1885–1940, suggesting that the reduction of production frictions accounted for most of the structural change in the Russian economy during this period. Yet, an open question remains, as to what factors and/or policies were responsible for the reduction in frictions. Blasco-Piles and Tadei first reconstruct an annual time series of production frictions in 1885–1940 by integrating Cheremukhin et al. (2017). The authors apply the wedge accounting methodology of Cheremukhin et al. (2017) to data from primary and secondary sources, extending the examination of frictions to the period from WWI to 1927, which was missing in the previous study. Additionally, new time series on literacy rates, inflation, and population have also been constructed and/or improved. The authors then use basic regression analysis to explore the relationship between production frictions and different periods, which in the Russian/Soviet case are clearly associated with different economic and industrial policies. Additionally, the authors check for discontinuities in the production-wedge time series. Both approaches lead to the argument that production frictions were rapidly reduced during Stalin’s era (1928–1940). The authors maintain that Stalin’s five-year plans reduced frictions due to their high level of investments, lax provision of bank credit to the heavy industry state-owned enterprises, and high production objectives.
The third article of the special issue, by Paschalis Arvanitidis and Charalampos Sofiadis (Arvanitidis and Sofiadis 2023), applies Ostrom’s analytical framework to the history of the Byzantine Monastery. Studies on the so-called “Commons”, i.e., institutional structures that communities develop to self-manage their common resources, have been blossoming in recent decades. Yet, most scholars focus on contemporary cases that are generally short-lived. Instead, the authors argue, studying historical Commons can enrich our understanding of how such institutions emerged and evolved over time, and provide inspiration for contemporary problems affecting similar social contexts and communities. The Byzantine monastic institution is analysed through the lens of Ostrom’s framework. The paper uses historical-institutional analysis to outline the structure of the institution, as well as Ostrom’s design principles (whether and how a community implements rules to avoid the “tragedy of the Commons”), to explore if the Byzantine monastery can be seen as an early example – and a long-lasting one – of a Common. Arvanitidis and Sofiadis argue that the Byzantine Monastery has been successful over time due to its structure and organization as a Common to manage limited resources. The community, which inhabited the monasteries, developed a credible governance structure that enabled it to successfully self-manage their resources and established the Byzantine Monastery as an important economic, social, cultural and spiritual institution which survived for many centuries.

The fourth article, by Zhongjin Li (Li 2023), studies three large East Asian economies – Japan, South Korea, and (to a lesser extent) China – that experienced an “economic miracle” based on sustained industrialization in the post WWII era. The three East-Asian countries have already received significant attention within the literature on economic development, due to their success and rapid growth based on a mix of market forces and state intervention. Yet, as Li argues in her article, most accounts of sustained economic growth in Asia do not consider an important aspect, namely distributional and class conflict. In an effort to fill this gap, Li’s article offers a labour-centred account of the fundamental class underpinnings upon which market and state forces operate to generate and share productivity gains. She relies on a labor-regime framework and applies historical analysis, statistical evidence, and econometrics to investigate and re-interpret the connection between class dynamics and distributional outcomes, as well as the way that this affected development and growth. The results show a close association between economic performance and the evolution of class compromise during capitalist development in East Asia, stressing the importance of class dynamics for a better understanding of economic growth and development in a capitalist context – beyond analyses of the same type of the West.
Acknowledgements

We wish to thank Michelangelo Vasta for providing invaluable comments on a draft of this article.

Author contribution statement


References


Introducción: teoría y econometría en el análisis histórico

Resumen

Este artículo supone una introducción al presente número monográfico de la Revista de Historia Industrial – Industrial History Review sobre «Teoría y econometría en el análisis histórico». En primer lugar, revisa brevemente la bibliografía sobre la evolución de la historia económica desde la llamada revolución cliométrica, es decir, el uso creciente de la teoría para enmarcar la evidencia histórica, así como de técnicas cuantitativas, sobre todo para estudiar la causalidad. En segundo lugar, proporciona una descripción general de cómo una posible revolución en el tipo y uso de datos, impulsada por el progreso tecnológico en la recopilación y el procesamiento de información histórica, puede estar afectando a la historia económica. Por último, el artículo proporciona un breve resumen de los trabajos incluidos en este número especial, los cuales ponen de relieve cómo la teoría económica y los nuevos análisis cuantitativos pueden usarse para explorar una variedad de temas dentro de la disciplina de la historia económica, que abarcan diferentes países y periodos de tiempo.

Palabras clave: cliometría, cliodinámica, historia económica, teoría, análisis histórico, datos, revolución.

Códigos JEL: A12, B25, B41, N01.