NBER WORKING PAPER SERIES

UNVEILING SPECIALIZATION TRENDS IN ECONOMICS RESEARCH: A LARGE-SCALE STUDY USING NATURAL LANGUAGE PROCESSING AND CITATION ANALYSIS

Sebastian Galiani
Ramiro H. Gálvez
Ian Nachman

Working Paper 31295
http://www.nber.org/papers/w31295

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023

We did not received any funding to do this research. We do not have anything to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Unveiling Specialization Trends in Economics Research: A Large-Scale Study Using Natural Language Processing and Citation Analysis
Sebastian Galiani, Ramiro H. Gálvez, and Ian Nachman
NBER Working Paper No. 31295
June 2023
JEL No. A1

ABSTRACT

This article presents a comprehensive analysis of trends in the publication and citation of economics scholarly research, with a focus on specialization within fields of economics research (i.e., applied, applied theory, econometrics methods, and theory). We collected detailed data on 24,273 articles published from 1970 to 2016 in highly regarded general research economics journals. We then used state-of-the-art machine learning and natural language processing techniques to further enrich the collected data. Our findings reveal significant disparities in article content and citations across fields of economics research. The analysis indicates growing specialization trends in theory and econometric methods. In contrast, applied papers are covering a wider range of topics and receiving an increasing proportion of extramural citations over time. By 2016, applied ranked among the most or second most cited field by any other field of economics research. These patterns are consistent with applied papers becoming more multidisciplinary. Applied theory articles have also demonstrated a growing breadth of topics covered (similar to applied articles); however, this has not been accompanied by an increase in extramural citations or in the share of citations received from other fields of economics research (as observed with theory articles). This makes it challenging to determine their specialization status.

Sebastian Galiani
Department of Economics
University of Maryland
3105 Tydings Hall
College Park, MD 20742
and NBER
sgaliani@umd.edu

Ramiro H. Gálvez
Universidad Torcuato Di Tella
Av. Figueroa Alcorta 7350, (C1428BIJ)
Buenos Aires
Argentina
rgalvez@utdt.edu

Ian Nachman
Brown University
ian_nachman@brown.edu

Unveiling Specialization Trends in Economics Research: A Large-Scale Study Using Natural Language Processing and Citation Analysis

Sebastian Galiani[†]
University of Maryland

Ramiro H. Gálvez[‡]
Universidad Torcuato Di Tella

Ian Nachman[§]
Brown University

May 24, 2023

Abstract

This article presents a comprehensive analysis of trends in the publication and citation of economics scholarly research, with a focus on specialization within fields of economics research (i.e., applied, applied theory, econometrics methods, and theory). We collected detailed data on 24,273 articles published from 1970 to 2016 in highly regarded general research economics journals. We then used state-of-the-art machine learning and natural language processing techniques to further enrich the collected data. Our findings reveal significant disparities in article content and citations across fields of economics research. The analysis indicates growing specialization trends in theory and econometric methods. These articles have shown a narrowing breadth of topics covered, and their share of extramural citations and citations from other fields of economics research has either remained stable or decreased. In contrast, applied papers are covering a wider range of topics and receiving an increasing proportion of extramural citations over time, particularly from disciplines such as medicine, psychology, law, and education. In addition, citations to applied papers coming from other fields of economics research have increased during the period analyzed. By 2016, applied ranked among the most or second most cited field by any other field of economics research. These patterns are consistent with applied papers becoming more multidisciplinary. Applied theory articles have also demonstrated a growing breadth of topics covered (similar to applied articles); however, this has not been accompanied by an increase in extramural citations or in the share of citations received from other fields of economics research (as observed with theory articles). This makes it challenging to determine their specialization status. To our knowledge, this study represents the first comprehensive attempt to integrate content analysis with citation analysis to document specialization trends within fields of economics research, or in the publication and citation of economics scholarly research in general. (JEL A11, A14)

1. Introduction

Knowledge is commonly divided into fields of study or academic disciplines (e.g., economics, psychology, mathematics, etc., see Krishnan, 2009). These are typically defined and circumscribed by the journals that publish their research, their learned societies, and the academic departments to which their practitioners belong. Usually, a field of study focuses on a series of research topics. However, defining a field of study solely by the research topics it addresses is inadequate, as, in many cases, the same research topic is addressed by multiple fields of study by means of different research strategies/tools. A well-documented fact is that, as time passes by and knowledge accumulates, fields of study tend to become more specialized, to address increasingly complex research questions and to develop new, tailored research tools (see Ramaley, 1930; Ziman, 1987; Wray, 2005; *inter alia*). Although specialization is believed to lead to deeper and more precise knowledge, it is also believed to have created barriers to communication and collaboration across fields of study (see, for example, Becker and Murphy, 1992; Walsh and Maloney, 2007; Anderson and Richards-Shubik, 2022).[1]

[†] sgaliani@umd.edu; [‡] rgalvez@utdt.edu; [§] ian_nachman@brown.edu

[1] William Dunham, in his book "*The mathematical universe*" (Dunham, 1994), presents the reader with an exercise that illustrates this trend. He notes that any mathematician asked to name history's three or four most influential mathematicians would surely include Sir Isaac Newton in their list. Similarly, any physicist asked to name history's three or four most influential physicists would also include Newton. Dunham concludes that, although this is extraordinarily rare, it occurs in part because Newton worked at a time before "*unscalable walls were erected between disciplines*." He also notes that, at the time, subjects such as optics, astronomy, and mechanics were in fact treated as branches of mathematics.

Within a field of study, knowledge is also typically divided into subfields.[2] Subfields of study usually have their own specialized journals and learned societies, and researchers within a subfield tend to focus their research on a subset of the topics addressed by the parent field. However, another common division within fields of study is that of dividing the field into a theoretical branch and an experimental/applied branch. As an illustrative example, physics is commonly divided into theoretical physics and experimental physics (see Duhem, 1976), where theoretical physicists use advanced mathematical models to develop theories that explain physical systems and experimental physicists design and conduct experiments to test these theories. Similar divisions are observed in other fields of study, such as theoretical and empirical/experimental psychology, mathematical/theoretical and experimental biology, theoretical and experimental linguistics, and so on. Ideally, a feedback loop should exist between the two branches, the theoretical one providing new theories to be tested, and the empirical one providing novel data useful in the development, refusal, or validation of these theories (see Putnam, 1979; Reeves et al., 2008).

Economics is no exception, and a clear division between theoretical and applied economics is believed to exist (see Backhouse and Cherrier, 2014, Hamermesh 2018; Joshua Angrist et al. 2017; *inter alia*). Specialized journals exist for both theory and applied economics, applied/theory oriented postgraduate programs exist, and, some universities even have dedicated applied economics departments. What is less clear is how these two branches interact with each other. Do they feedback each other? Or, as it is believed to be the case across fields of study, they are becoming so specialized that collaboration is being precluded? Focusing on what we refer to as "fields of economics research" (*applied*, *applied theory*, *econometric methods*, and *theory*), we empirically address this issue by analyzing long-term trends in the contents of, and the citations received and generated by, a large and representative corpus of economics research articles. Concretely, we look for patterns that indicate specialization within a field of economics research, such as narrowing the topics it covers in a way that other fields do not, and patterns suggesting a decrease in citations from outside the field. We define a field of economics research based on the methodological techniques/strategies used by its articles to address research topics/questions. See Section 2.3 for a discussion and details on the criteria we use to classify papers into fields of economics research.[3]

The research strategy we implement in this article is as follows. We first collect a large sample of economics research articles published in top outlets. The sample covers a wide and representative set of research topics/strategies and spans the period 1970-2016. Then, for each article in our sample we collect extensive detailed data (e.g., metadata — authors, year of publication, etc.; wordcounts; details on the articles citing them; details on the articles they cite; and so on). Having collected these data, we enhance/enrich it by making use of modern machine learning (ML) and natural language processing (NLP) techniques. For example, 1) based on the articles' wordcounts and on a set hand-labelled field of economics research tags obtained from previous studies, we identify the field of economics research to which each article in our sample belongs; 2) based solely on the articles' wordcounts, we identify the topics addressed by each article in our dataset; and, 3) based on the titles and abstracts transcripts of the economics research articles citing, or being cited by, our sample, we identify to the field of economics research to which these citations/references belong. The resulting data enable us to examine previously undocumented trends characterizing specialization in economics. For example, up to what degree applied papers cite/reference theory articles, the degree to which different fields of economics research have concentrated their research efforts on a few topics, which topics have gained more presence in theory or applied papers in the last decades, and so on.

[2] For example, medicine is divided into anatomy, physiology, pathology, etc.; chemistry is divided into analytical chemistry, organic chemistry, inorganic chemistry, etc.; economics is divided into labor economics, behavioral economics, economic development; etc.

[3] As stated in Hamermesh (2013), "*subject does not imply method*," as a given economics research topic (e.g., corruption) is commonly addressed by means of different research strategies (e.g., by running a lab experiment or by means of a theoretical model). In fact, it could easily be argued that researchers require less training to migrate from one research topic to another while using the same research methodology than to study the same topic by means of different research methodologies.

The division of a field into empirical and theoretical branches is a widely recognized and extensively studied phenomenon across multiple disciplines (see, Poincaré, 1902; Dirac, 1940; Koch, 1973; Harrow, 1979; Longo and Soto, 2016; Phillips et al., 2021; *inter alia*). Previous papers within the field of economics have also examined differences along this dimension. Hamermesh (2013) provides data on articles published in the top three general economics journals for one year in each decade from the 1960s to the 2010s. The author documents that these journals are publishing fewer papers that represent pure theory, fewer papers based on publicly available datasets, and more empirical studies based on data collected by the author(s) or on laboratory or field experiments. In Hamermesh (2018) a field of economics research tag is assigned to 439 articles published in 1974-75 and to 497 articles published in 2007-08. The author documents that empirical and experimental articles receive more citations than economic theory or econometric theory articles do. Chiappori and Levitt (2003) explore whether theoretical economic research succeeds in influencing the path of empirical microeconomic research. They examine a dataset consisting of empirical microeconomic papers published between 1999 and 2001. They find that the set of theoretical papers cited as a primary motivation for empirical research projects is surprisingly diverse and that very few theoretical papers have much of an influence on applied microeconomic papers. Biddle and Hamermesh (2017) code the content of all applied microeconomic articles published in the top economics journals in the years 1951-1955, 1974-1975, and 2007-2008. They identify three separate stages in the post-World War II history of applied microeconomic research (a nonmathematical period, a period of consensus, and a period referred to as the experimentalist paradigm). They also report a partial abandonment of theory by applied microeconomists. Backhouse and Cherrier (2014) examine the historical winners of the John Bates Clark medal. They conclude that economics has become increasingly applied over time, with applied work now being accorded higher status compared to pure theory. Anauati, Galiani, and Gálvez (2016) characterize how life cycles in yearly citations differ across four fields of economic research (applied, applied theory, econometric methods and theory — the same categories we use in this article). They assign a field of economics research tag to 9,672 economics research articles published in top journals between 1970 and 2000 and collect detailed citation data for them. They document strong differences in citation patterns across fields of economics research. For example, applied and applied theory articles receive higher numbers of citations per year and have the longer citation lifespans. Using the same fields of economics research categories, Anauati, Galiani, and Gálvez (2020) study how citation patterns differ between journal tiers in economics. They assign fields of economics research tags to 6,083 articles published across different journal tiers. Among their findings, they report that the differences in citation performance across tiers vary depending on the fields of economics research covered by their articles. In this article, we use data from Anauati, Galiani, and Gálvez (2016) and from Anauati, Galiani, and Gálvez (2020) to train ML/NLP models that assign a field of economics research tag to a much larger sample of economics research articles (see Section 2.3 and Section 3).

The two papers most closely related to ours are Angrist et al. (2017) and Angrist et al. (2020). In both articles, the authors utilize ML/NLP techniques to assign specific field of economics research tags to a large sample of economic research articles. Both articles examine three distinct fields of economics research: *empirical*, *theory*, and *econometrics*. They use the term "style" to refer to these fields. In Angrist et al. (2017), the authors examine the purported shift in economics from theoretical to empirical research. They document that the growth in empirical work reflects a shift within rather than across "fields." The authors use the term field to refer to a collection of research topics, for example, microeconomics, labor, macroeconomics, etc. They also document that empirical work is now cited more often than theoretical work. In Angrist et al. (2020), the authors look at the influence of economic research through the lens of *extramural citations* (i.e., citations received from fields of study other than economics). The authors document a clear rise in the extramural influence of economics research, along with a growing tendency for economics to reference and draw upon other social sciences. They observe that a significant portion of the increase in economics' external influence can be attributed to the rise in citations to empirical work. Additionally, they document an increase in the proportion of citations attributed to empirical papers within the field of economics.

The present paper differs from previous studies in two important ways. First, its focus is placed primarily on documenting specialization trends within fields of economics research. In doing so, we effectively document patterns that have been previously studied, such as the shares of extramural citations received by different fields of economics research (Angrist et al., 2020). However, we also go beyond existing literature by uncovering previously unexplored trends, providing a comprehensive overview of the historical evolution of economics research. Second, our paper introduces several methodological innovations. On the one hand, we analyze new data sources that provide a higher level of granularity compared to those examined in previous studies. This enables us to study a larger and more representative corpus, spanning a longer period of time than most previous literature.[4] Additionally, we collect more extensive data on each analyzed article. On the other hand, the collected data allows us to utilize ML/NLP techniques that have not been employed in previous literature studying corpora of economics research articles. For instance, to the best of our knowledge, this study represents the first application of topic analysis on such a vast and comprehensive corpus of economics research articles.

Our analysis reveals significant disparities in the content, citations, and references of articles across fields of economics research. Certain fields demonstrate a growing trend towards specialization, while others exhibit contrasting patterns. Specifically, theory and econometric methods have shown a narrowing focus on specific research topics since the 1990s, indicating a tendency towards specialization. Theory articles have experienced a significant increase in topics related to formal mathematical proofs and game theory, while econometric methods articles have shown a pronounced rise in topics related to computational statistics, estimators' asymptotic properties, and estimators' bounds. In comparison to applied and applied theory, these fields receive fewer extramural citations and citations from other fields of economics research, suggesting specialization. In contrast, applied papers have expanded their coverage to include diverse topics. Applied articles have seen a pronounced rise in topics related to impact analysis, causal analysis, and experimental economics. Over time, applied articles began to receive a higher proportion of citations from external sources, especially from disciplines such as medicine, psychology, law, and to a certain extent, education. By 2016, applied ranked among the most or second most cited field by any other field of economics research. Overall, these patterns indicate a shift toward multidisciplinary. The case of applied theory articles is less conclusive. While they cover a broader range of topics (similar to applied papers), there has been no significant increase in extramural citations or citations from other fields of economics research (as observed with theory articles). In fact, applied theory articles receive the smallest share of extramural citations. These opposing patterns make it challenging to determine their specialization status. Overall, the findings indicate a trend of specialization in theory and econometric methods, a move towards multidisciplinarity in applied papers, and a nuanced situation in applied theory.

The rest of the article is structured as follows. Section 2 provides a detailed description of the data sources utilized in this study. In Section 3, we explain the filtering, processing and enrichment of the collected data through the application of ML/NLP techniques, leading to the creation of our final datasets. Section 4 presents the main findings of our analysis. Finally, Section 5 concludes by summarizing our results and discussing the practical implications of our analysis.

2. Data sources

We utilize data from three sources: 1) *Constellate*, 2) *Semantic Scholar*, and 3) a set of fields of economics research tags taken from Anauati, Galiani and Gálvez (2016) and Anauati, Galiani and Gálvez (2020). Below, we provide further details on each of these sources.

[4] For the case of styles (what we refer to as fields of economics research), Angrist et al. (2017) and Angrist et al. (2020) study the period from 1980 to 2015, which is slightly shorter than the period we analyze (1970-2016). The remaining literature cited either studies shorter periods or investigates larger periods but with significant sampling from the population of economics articles published in the journals they analyze.

2.1. Constellate

Launched in 2021, Constellate is a text analytics service provided by ITHAKA (the not-for-profit organization managing both JSTOR and Portico).[5] Constellate includes an online dataset builder which allows to retrieve data from articles satisfying a series of criteria (e.g., articles that have been published in a given journal, that have been published in a given year, etc.). The downloaded data includes metadata for each article that meets the criteria, such as title, digital object identifier (DOI), author names, publication date, journal of publication, and more. Importantly, it also includes the counts of unigrams, bigrams, and trigrams associated with each article retrieved.[6]

Constellate covers only a subset of all published economics research journals. For example, data for articles from the journal *Economic Theory* cannot be downloaded from Constellate. Taking this into account, we opted to fetch Constellate data for as many general research journals as were studied in Anauati, Galiani and Gálvez (2016) and Anauati, Galiani and Gálvez (2020). That is, we downloaded data for all articles published from 1970 up to and including 2016 in the so-called economics *Top 5* journals (*The American Economic Review*, *Journal of Political Economy*, *The Quarterly Journal of Economics*, *Econometrica*, and *The Review of Economic Studies*) and for all articles published from 1970 up to and including 2016 in a set of *non-Top 5* general research journals (*The Economic Journal*, *International Economic Review*, *Economic Inquiry*, and *The Review of Economics and Statistics*).[7] Taken together, these journals extensively cover all fields of economics research and have been the subject of study in numerous articles similar in scope to ours (e.g., Heckman and Moktan, 2020; Angrist et al., 2020; Card and DellaVigna, 2013; Hamermesh, 2018, etc).

Solely for training the NLP models described in Section 3.1 and Section 3.2, we downloaded Constellate data for articles published from 1970 to 2016 in *The Journal of Law & Economics* and *The RAND Journal of Economics*. These are the only "top field" journals considered in Anauati, Galiani and Gálvez (2020) that are available for download from Constellate throughout the entire period.

After downloading this data, we applied a series of filters. First, we retained documents that had the tag "research-article" or "article" in the "docSubType" field, removing those classified as miscellaneous by Constellate. Second, we dropped all documents for which the field “creator” (which lists the articles’ authors) was empty or missing. Third, if two articles shared titles and one was retrieved from JSTOR while the other was from Portico, we retained only the JSTOR version. Fourth, we removed certain articles identified as not being proper research articles (e.g., replies, rejoinders, errata, editor's reports, software reviews, etc.) by analyzing their titles. Finally, we discarded documents with missing "title" fields. After applying these filters, the Constellate data consisted of 34,623 documents.

2.2. Semantic Scholar

As we will describe in detail in Section 3.2, we downloaded Semantic Scholar data for all articles included in the Constellate data. Semantic Scholar is an artificial intelligence powered research tool for scientific literature developed by the Allen Institute for AI.[8] Launched in 2015, the tool provides an academic search engine (accessible through an API), and, by early 2023, indexed more than 207 million published scientific research articles, books, and articles’ preprints (see Matthews, 2021). The Semantic

[5] https://constellate.org/

[6] Unigram count stands for the time a given word/token appears in a document. Bigram counts stands for the times a given sequence of two words/tokens (e.g., "your homework") appears in a document. Trigram counts is the equivalent, but for sequences of three words (e.g., " turn your homework"). See Jurafsky and Martin (2009).

[7] We downloaded data up to and including 2016, as, at the moment of download, Constellate did not include data on newer articles.

[8] https://allenai.org/

Scholar's corpus includes documents from all fields of research and is regarded as comparable in coverage to that of Scopus and the Web of Science (see, Martín-Martín et al., 2021).[9]

After conducting a search, Semantic Scholar provides various data, including metadata associated with all the retrieved articles (title, publication date, article identifiers, authors, field of study tags, etc.), their abstracts, comprehensive data on their references (i.e., the articles cited by the retrieved articles), and detailed data on all articles that cite the retrieved articles. The detailed data of reference and citing articles includes, among other information, the years of publication, abstracts, and fields of study tags.

The field of study tags are labels that assign a given article to one or more of the following 23 fields of study: 1) economics, 2) history, 3) philosophy, 4) computer science, 5) business, 6) medicine, 7) physics, 8) political science, 9) mathematics, 10) psychology, 11) sociology, 12) geology, 13) environmental science, 14) law, 15) biology, 16) engineering, 17) education, 18) art, 19) geography, 20) agricultural and food sciences, 21) materials science, 22) chemistry, and 23) linguistics.[10] Importantly, in this study, we consider an article to be an economics article if "economics" is listed among its Semantic Scholar fields of study tags, regardless of whether any other field is also listed.

2.3. Fields of economics research hand-labelled tags

In Anauati, Galiani, and Gálvez (2016), a total of 9,672 articles published in the Top 5 journals from 1970 to 2000 were manually tagged as belonging to one of the following four fields of economics research: 1) applied, 2) applied theory, 3) econometric methods, and 4) theory. The definition of field of economics research used in that article aimed to capture the research strategy employed and the skills required when writing an economics article. The articles tagged were sampled from EconLit[11] and the criteria used to assign a paper to a field were as follows: 1) *Applied* articles have an empirical or applied motivation. They rely on the use of econometric or statistical methods as a basis for analyzing empirical data, although they may deal with simple models that serve as a theoretical framework for the analysis. The category also includes papers which do not use sophisticated econometric methods, but do use descriptive statistics to analyze, for example, given features of an economy and in which the empirical section figures as the central element. 2) *Applied theory* articles develop a theoretical model to explain a fact; the empirical analysis is not the most important feature of the paper, but a supplement. In these papers, the use of econometric or statistical analyses is limited, although they may use simulations (even with empirical data) or refine other techniques to test the implications of the models.[12] 3) *Econometric methods* articles develop econometric or statistical methodologies. They also include papers that develop methodologies for collecting data and that address issues of identification, data aggregation, or optimization techniques. 4) *Theory* articles do not contain an empirical fact section; they usually approach a topic by modeling and by making extensive use of formal mathematics and logic. They may include a numerical example or a simple model calibration with theoretical data to illustrate the proposed model or analyze its comparative statics. Further information on the classification criteria can be found in Anauati, Galiani, and Gálvez (2016)

In addition to the tags assigned to the articles mentioned earlier, we also utilize the field of economics research tags assigned to 549 articles published in the Top 5 journals in 2005 and 2010. These tags were initially assigned manually during the early stages of the tagging process conducted in Anauati, Galiani, and Gálvez (2016), but were ultimately not used in their final analysis.

[9] This comes mainly from the fact that Semantic Scholar incorporated data from Microsoft Academic available via the Microsoft Academic Graph (see, Martín-Martín et al., 2021).
[10] For details on how these tags are assigned, see https://blog.allenai.org/9d2f641949e5.
[11] https://www.aeaweb.org/econlit/.
[12] Note that according to the criteria followed in Angrist et al. (2020), these articles would be tagged as "empirical" articles, as the authors claim that the empirical category should be understood as "not purely theoretical."

In Anauati, Galiani, and Gálvez (2020), the same criteria for assigning articles to fields of economics research were employed to tag a series of additional articles published from 1992 to 1996 in non-Top 5 general research journals and top field journals. We utilize these tags for all articles published in the journals analyzed in Anauati, Galiani, and Gálvez (2020) that are also available in Constellate throughout the entire analyzed period. This adds a total of 1,681 article tags to our dataset.

After excluding seven articles identified as irrelevant or duplicates, we utilize a total of 11,895 fields of economics research tags. We will refer to these hand-labeled tags as the *training tags*, and the collection of these articles as the *manually tagged dataset*. Table 1 displays the distribution of training tags across the journals included in the manually tagged dataset.

Table 1. Training tags distribution across journals in the manually tagged dataset

	Applied	Applied theory	Econometric methods	Theory
American Economic Review	990 (37%)	359 (13%)	47 (2%)	1305 (48%)
Econometrica	172 (7%)	155 (6%)	786 (33%)	1295 (54%)
Economic Inquiry	141 (64%)	41 (19%)	0 (0%)	37 (17%)
Economic Journal	181 (50%)	46 (13%)	23 (6%)	109 (30%)
International Economic Review	19 (7%)	34 (13%)	22 (8%)	186 (71%)
Journal of Labor Economics	94 (70%)	17 (13%)	1 (1%)	22 (16%)
Journal of Law & Economics	85 (86%)	5 (5%)	0 (0%)	9 (9%)
Journal of Political Economy	844 (40%)	286 (14%)	14 (1%)	941 (45%)
Quarterly Journal of Economics	516 (34%)	148 (10%)	11 (1%)	824 (55%)
RAND Journal of Economics	75 (38%)	11 (6%)	2 (1%)	108 (55%)
Review of Economic Studies	128 (8%)	129 (8%)	140 (9%)	1125 (74%)
Review of Economics and Statistics	320 (78%)	21 (5%)	66 (16%)	5 (1%)
Total	3565 (30%)	1252 (11%)	1112 (9%)	5966 (50%)

3. Data construction

This section serves two purposes. First, it provides a detailed walkthrough of how the data presented in Section 2 is processed, enriched, and merged to build our final datasets. Second, it provides a brief introduction to the NLP techniques used in this article, including both traditional and state-of-the-art techniques.

3.1. Assigning a field of economics research tags to all of the Constellate articles

After downloading the Constellate detailed data (Section 2.1) and collecting hand-labelled fields of economics research tags (Section 2.3), the first step in building our final dataset is to assign each Constellate article a field of economics research tag. Since articles with manually-assigned tags represent a subset of the Constellate data, the process of assigning tags to all Constellate articles is more complex than a straightforward merge of the two datasets. In this section, we provide a detailed explanation of the approach we employed to predict field of economics research tags for articles without a training tag.

We approached the problem as a multilabel classification task (Hastie, Tibshirani, and Friedman 2009), where the training data consisted of Constellate articles that already had a manually assigned field of economics research tag. To create this training data, we needed to merge the Constellate data with the field of economics research tags. However, merging the two datasets was not straightforward due to differences in DOIs between Constellate articles and the articles in the manually tagged dataset. This discrepancy arises because JSTOR and Portico commonly assign a new DOI to each article, which is different from the one originally assigned by the publishers.

To overcome this challenge, we implemented the following strategy. First, for every article i in the manually tagged dataset, we identified all Constellate articles published in the same journal and year as i. Second, we calculated the Levenshtein distance between the titles of the identified

Constellate articles and article i's title.[13] Finally, we assigned the field of economics research tag from article i to the Constellate article with the lowest Levenshtein distance. As a result of this process, we obtained a training dataset consisting of 11,624 Constellate articles for which we were able to accurately assign a field of economics research tag manually.

From the training data, we created a *document-term matrix*, which is a matrix where each row represents an article, each column corresponds to a word/token, and the cell values indicate the frequency of each word in each document. The word frequencies were obtained from the Constellate unigram counts. Before incorporating these counts into the document-term matrix, we performed several clean-up procedures on the Constellate unigram counts. First, we converted all words to lowercase. Second, we removed leading and trailing whitespaces. Third, we eliminated non-alphanumeric characters at the beginning and end of words. We also included columns in the document-term matrix to account for words appearing in the titles of the articles.[14] To calculate word frequencies, we applied the tf-idf transformation and excluded words that did not appear in at least 0.5% of all articles, as well as words that appeared in more than 80% of all articles.[15]

We chose an L_2 regularized multinomial logistic regression as our classifier (for further details see Hastie, Tibshirani, and Friedman, 2009 and Gentzkow, Kelly, and Taddy, 2019). This approach, utilizing a linear classifier with a document-term matrix as input, is widely used in document classification tasks (see, for example, Gentzkow, Kelly, and Taddy, 2019).[16] To improve the accuracy of the classifier, we applied inverse proportional weighting to all training observations based on the class frequencies in the input data.[17] Our final model had an inverse regularization strength equal to 1.5 (C) and a maximum number of iterations equal to 150.[18] We determined these values through a cross-validation exercise, using 80% of the data for training and 20% for validation (n=9299 and n=2325, respectively). Table 2 presents the main performance metrics obtained in the validation set, while Table 3 shows the confusion matrix, providing additional information about the predictions in the validation set.

Table 2. Validation predictive performance for fields of economics research tags predicted from Constellate data

	Precision	Recall	F_1-Score	n
Applied	0.86	0.88	0.87	677
Applied theory	0.63	0.59	0.61	277
Econometric methods	0.81	0.82	0.82	190
Theory	0.94	0.94	0.94	1181
All (weighted average)	0.87	0.87	0.87	2325

Notes. Precision is equal to the proportion of all observations *predicted* as belonging to a given class that were correctly classified. Recall is equal to the proportion of all observations *effectively belonging* to a given class that were correctly classified. The F_1-Score is equal to the harmonic mean between precision and recall.

[13] The Levenshtein distance is a metric which measures the distance between two texts as the minimum number of single-character edits required to change one word into the other. See Jurafsky and Martin (2009) for more details. Before doing this procedure, we ran few clean-up procedures to both the Constellate and EconLit titles (for example, lowercasing them, striping their leading and trailing whitespaces, removing all sequences of words inside brackets, etc.)

[14] We prefixed the string "*title_*" to each of these words. This was done to distinguish the occurrences of words in the titles of the articles from those in the bodies of the articles.

[15] The tf-idf transformation (short for term frequency-inverse document frequency transformation) is a numerical transformation in which the presence/importance given to a word in a document in a given corpus increases according to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word. It is commonly used when representing a corpus by means of a document-term matrix. See Jurafsky and Martin (2009) for more details.

[16] As Constellate data only includes unigram, bigram, and trigram counts and does not provide full-text transcriptions, we were unable to utilize more advanced approaches such as BERT-like models with the available data (See Section 3.2).

[17] For further details, see the "balanced" option of the parameter "class_weight" in scikit-learn's logistic regression documentation (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html).

[18] As detailed in Hastie, Tibshirani, and Friedman (2009) and Gentzkow, Kelly, and Taddy (2019), L_2 penalized models optimize a function where a regularization component is added to the loss function. Typically, this component has the following structure $\lambda \sum_{j=1}^{q} \beta_j^2$, where β_j stands for the regression coefficient associated to variable j, q stands for the total number of predictors, and λ for the regularization parameter (which is typically set by means of a cross-validation exercise). Inverse regularization strength (C) is defined as $1/\lambda$.

Table 3. Confusion matrix for fields of economics research tags predicted from Constellate data, validation set

	Applied	Applied theory	Econometric methods	Theory
Applied	598	59	8	12
Applied theory	62	164	12	39
Econometric methods	11	8	156	15
Theory	22	29	16	1114

Notes. Rows represent the actual class and columns represent the predicted class.

Our classifier achieves a competitive performance, having an accuracy equal to 0.87 (the trace of Table 3 divided by the sum of all its values) and a weighted average F_1-Score also equal to 0.87. As a benchmark, in Angrist et al. (2020), the authors report achieving an accuracy equal to 0.80 in a similar exercise.[19] Both Table 2 and Table 3 demonstrate the exceptional performance of our classifier in predicting applied, econometric methods, and theory tags. However, the performance declines when predicting applied theory labels. Table 2 highlights that the most frequent misclassification made by the classifier is predicting applied theory articles as applied ones.

After validating our predictive model, we proceeded to retrain it using the identical configuration as described above, but this time using the entire training set. Subsequently, we used the retrained model to predict the fields of economics research tags for all Constellate articles that were not part of the manually tagged dataset (n=22,999). However, for the articles included in the manually tagged dataset, we retained the manual tags and did not replace them with predicted ones. This resulted in a total of 34,623 articles with assigned fields of economics research tags.

3.2. Assigning a field of economics research tag to all of the articles' references and citations

After assigning a field of economics research tag to each article in the Constellate corpus, we proceeded to download data on their detailed references and citations from Semantic Scholar. However, matching the data from Semantic Scholar to Constellate is not straightforward due to differences in DOIs between the two sources. To ensure a reliable merge, we took the following steps. First, we conducted a title-based search in the Semantic Scholar data for each of the 34,623 articles in the Constellate dataset and obtained the API response (we were able to download data for all but 419 articles). Second, to ensure the accuracy of the merge, we implemented a series of filters. We dropped articles for which author information was not available in either Constellate or Semantic Scholar. We removed merges that did not share at least one author surname in both data sources. We also excluded merges where the same Semantic Scholar article was matched to multiple Constellate articles, ensuring that each Semantic Scholar article had a single match in Constellate. Finally, we dropped merges where the Levenshtein distance between the Constellate and Semantic Scholar titles exceeded 15.

The process described above resulted in a total of 31,093 articles with detailed Constellate and Semantic Scholar data available. These articles were either referenced or cited by 2,040,265 different articles, out of which 1,330,590 were classified as economics research articles by Semantic Scholar. To examine how our sample of economics research articles is cited by or references articles from different fields of economics research, we predicted a field of economics research tag for each of the 1,330,590 economics articles referenced by or citing our final sample. Once again, we treated this problem as a multilabel classification task. However, due to the absence of word counts in Semantic Scholar data, we were unable to follow the exact procedure outlined in Section 3.1. Instead, we fine-tuned a BERT-like large language model (LLM, see Devlin et al., 2019) using the articles' titles (provided by Semantic Scholar), the articles' abstracts (if available from Semantic Scholar), and the hand-labelled fields of economics research tags described in Section 2.3. The training data for this task consisted of

[19] An important difference between our study and Angrist et al. (2020) is that they define "styles" differently. They label what we define as applied theory as "empirical." When we recode applied theory predicted and actual tags as "applied," our classifier achieves an accuracy of 0.93 on the validation set.

10,876 articles, which is a subset of the 11,624 articles used as the training set in Section 3.1. These 10,876 articles were chosen based on our high confidence in the availability and reliability of Semantic Scholar data.

Specifically, we fine-tuned a pre-trained SciBERT model (Beltagy, Lo, and Cohan, 2019) to adapt it for multiclass classification. Each input text consisted of the article title (preceded by the token "title: ") followed by a newline character, and then followed by the article abstract (preceded by the token "abstract: ").[20] Training instances were weighted inversely proportional to the class frequencies in the input data. The model configuration was determined through a validation exercise, where 80% (n=8700) of the observations were used for training different model setups and 20% (n=2176) were used for validation. Table 4 presents the main performance metrics obtained in the validation set, while Table 5 shows the confusion matrix, providing additional information about the predictions in the validation set.

Table 4. Validation predictive performance for fields of economics research tags predicted from Semantic Scholar data

	Precision	Recall	F_1-Score	n
Applied	0.83	0.77	0.80	675
Applied theory	0.55	0.24	0.34	238
Econometric methods	0.76	0.62	0.68	162
Theory	0.79	0.94	0.86	1101
All (weighted average)	0.77	0.79	0.77	2176

Notes. Precision is equal to the proportion of all observations *predicted* as belonging to a given class that were correctly classified. Recall is equal to the proportion of all observations that *effectively belong* to a given class that were correctly classified. The F_1-Score is equal to the harmonic mean between precision and recall.

Table 5. Confusion matrix for fields of economics research tags predicted from Semantic Scholar data, validation set

	Applied	Applied theory	Econometric Methods	Theory
Applied	521	30	9	115
Applied theory	70	58	11	99
Econometric methods	2	2	100	58
Theory	37	16	11	1037

Notes. Rows represent the actual class and columns represent the predicted class.

Both Table 4 and Table 5 demonstrate the classifier's strong performance in predicting applied and theory labels, while achieving lower accuracy for econometric methods and applied theory articles. The overall accuracy of the classifier in the validation set is 0.79, with a weighted average F1-Score of 0.77.[21]

Finally, we retrained the model using the same configuration as described above, but with the entire training set. With this final model, we predicted a field of economics research tag for each of the 1,330,590 articles that reference or cite our sample of articles. However, since the predictions in Section 3.1 performed better than the ones presented here (a clear example of better data beating better algorithms), we kept the previous predictions for any reference or citing article whose tag had already been predicted in Section 3.1.

Up to this point, we focused on assigning as accurate as possible fields of economics research tags to all articles published in our journals of interest, as well as to their references and to the articles citing them. In order to achieve this, we prioritized retaining articles with hand-labelled tags, even if they were not included in our final sample. After training the models as described above, we dropped two sets of articles that were not of interest to us. First, we dropped all articles published in *The Journal of*

[20] In case of an article not having its abstract available in Semantic Scholar, the input text consisted of consisted simply of the title (preceded by the token "title: ").
[21] When we categorize applied theory validation labels and predictions as applied, the model achieves an accuracy of 83%. Recall that Angrist et al. (2020) achieved an accuracy of 80% in a similar exercise.

Law & Economics and *The RAND Journal of Economics* (as this project focuses on general research economics journals, not field ones). Second, following previous studies such as Card and DellaVigna (2013) and Anauati, Galiani, and Gálvez, (2016), we filtered out all articles published in the *Papers and Proceedings* issues of *The American Economic Review*. This left a total of 24,273 economics research articles for which we collected detailed Constellate data, assigned a field of economics research tag, and collected detailed Semantic Scholar data. We refer to this set of articles as our *final sample*.[22]

3.3. Detecting latent topics in economics research articles

After constructing our final sample, our attention now turns to analyzing the content of the articles. For this purpose, we employ the latent Dirichlet allocation (LDA) generative model (Blei, Ng, and Jordan, 2003). As stated in Hu (2009), LDA is an unsupervised learning algorithm that assumes words hold significant semantic information and that documents discussing similar topics tend to use similar groups of words. It automatically identifies word groups that frequently co-occur within documents in a corpus and associates them with topics. This is accomplished by modeling documents as random mixtures over latent topics, each characterized by its own distribution over the vocabulary of words. To train the LDA model, we utilized Gensim's implementation of the LDA training algorithm (see Řehůřek and Sojka, 2010). We set the number of topics to 750 and performed 20 passes through the corpus during training.[23]

Before training the LDA algorithm, we performed several preprocessing steps to clean the input text obtained from Constellate's unigram counts. These procedures aimed to enhance the quality of the training data. First, we lemmatized all tokens, converting them to their lemma form to capture the core meaning of words (e.g., replacing *walking* by *walk*, see Jurafsky and Martin, 2009). Next, we removed tokens composed solely of numbers and those not containing any alphanumeric characters. Finally, we eliminated stopwords, which are commonly used function words with little or no semantic value (e.g., *the*, *this*, etc., see Jurafsky and Martin, 2009). These steps collectively improved the quality of the tokenized text, ensuring a more effective training process for the LDA model.

Having trained the LDA algorithm on our final sample, we obtained two datasets as the main outputs. The first dataset indicates the presence of words in topics, where topics tend to be formed by a small number of words. We denote the presence of word l in topic k as $w_{k,l}$, with $w_{k,l} \geq 0$ and $\sum_l w_{k,l} = 1$. The second dataset indicates the presence of topics in documents, where documents tend to cover a small number of topics. We denote the presence of topic k in document i as $t_{i,k}$, with $t_{i,k} \geq 0$ and $\sum_k t_{i,k} = 1$. We define an article i as "containing" topic k if the presence of topic k in article i (an output of the LDA model) ranks among the top ten topics with the highest presence in that article. This approach ensures that each article is associated with exactly ten topics.[24]

As a final step, we conducted a filtering process which excluded topics that were deemed non-meaningful (for example, a topic composed of the words *online*, *access*, *appendix*, *american*, and *vol* — which can be associated to instruction on how to access to online appendices) or that did not exhibit a clear thematic pattern (topics that were not predominantly represented by a few words). After this filtering process, we were left with a total of 695 meaningful topics.

Here are a few examples that demonstrate the patterns identified by the LDA model. After training the LDA model on our data, the algorithm identified a topic with the following five most important words: *crime* (16.54%), *criminal* (4.03%), *police* (3.80%), *arrest* (3.19%), and *rate* (2.31%) (in

---

[22] For each article referenced by or citing our final sample of articles, we collected detailed Semantic Scholar data; and for the subset of economics research articles referenced by or citing our final sample, we also assigned field of economics research tags.
[23] LDA has been widely used to uncover latent topics in research articles. Previous studies such as Hall, Jurafsky, and Manning (2008), Amami et al. (2016), Gálvez (2017) have successfully applied LDA in this context.
[24] We tested other thresholds and consistently obtained qualitatively similar results throughout our study.

decreasing order of importance — word presence in the topic in parentheses). This topic appears to be associated with research on the economics of crime. Three articles that strongly relate to this topic are "*Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error?*" (Levitt, 1998b), "*Reexamining criminal behavior: The importance of omitted variable bias*" (Mustard, 2003) and "*Juvenile crime and punishment*" (Levitt, 1998a). Another topic identified by the LDA model has *game* (17.82%), *player* (15.30%), *payoff* (5.12%), *equilibrium* (4.65%), and *play* (4.58%) as its five most important words (in decreasing order of importance). This topic appears to be related to research on game theory, and three articles that have high values for this topic are "*A folk theorem for repeated sequential games*" (Wen, 2002), "*The 'folk theorem' for repeated games with complete information*" (Wen 1994), and "*Extensive form reasoning in normal form games*" (Mailath, Samuelson, and Swinkels 1993).

4. Results

As mentioned above, our final dataset consists of 24,273 economics research articles. For each article, we collected detailed Constellate data, assigned a field of economics research tag, and collected detailed Semantic Scholar data. We also collected detailed Semantic Scholar data for each article referenced by or citing our final sample of articles. For the subset of economics research articles referenced by or citing our final sample, we also assigned a field of economics research tags. Finally, for each article in our final sample, we detected the topics they cover.

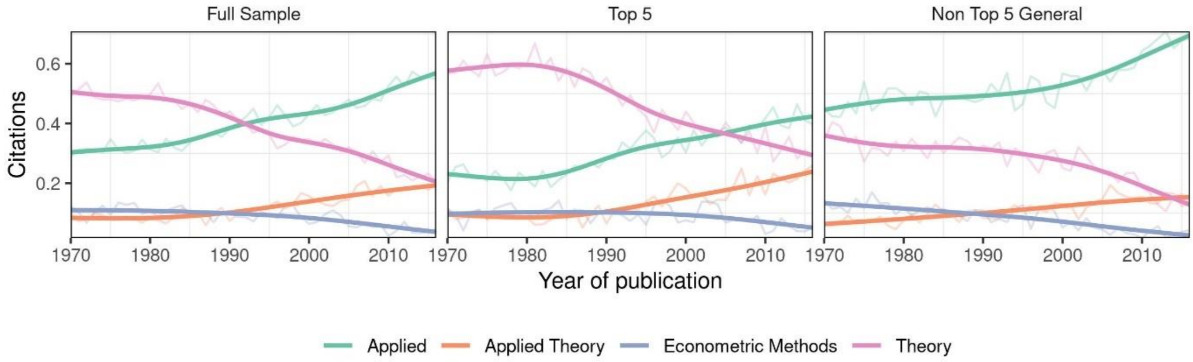
With this data available, we now study the level of variation in publishing practices within economics across different fields of research, with a particular focus on specialization within these fields. In general, we associate specialization with patterns that reveal a discipline is narrowing the topics it covers (where these topics are not being covered by other fields), as well as patterns suggesting the field is receiving less citation from outside fields. In Section 4.1 we analyze patterns and trends within the articles comprising our main sample. In Section 4.2 we investigate how citation patterns differ based on the field of study that cites, or is referenced by, economics articles (e.g., what do medicine articles cite from economics? Do these patterns differ across fields of economics research?). Third, in Section 4.3 we explore how economics articles reference other articles within the field (e.g., do applied articles tend cite other applied articles or they mostly cite theory ones?).

4.1. Aggregate trends across fields of economics research

In this section, we analyze aggregate trends across fields of economics research. Firstly, we examine the evolution of article shares from different fields of economics research over time. Next, we investigate trends in the citations received by articles from these fields. Finally, we explore aggregate trends in the topics covered by our final sample of articles.

In Figure 1, we present the annual distribution of articles across different fields of economics research published in the analyzed outlets from 1970 up to and including 2016. Specifically, for each field f and year y, we plot the proportion of articles published in year y that fall under field f. We depict this distribution for the entire sample, as well as separately for the Top 5 and the non-Top 5 general research outlets considered.

Figure 1. Trends in the yearly share of articles published by field of economics research



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009).

Figure 1 provides evidence of a notable shift in the publication landscape of general research journals in economics. The analysis reveals a substantial reduction in the share of theory articles, accompanied by a sharp increase in applied and applied theory articles. This finding is in line with previous studies by Hamermesh (2013), Panhans and Singleton (2017), Backhouse and Cherrier (2014), Hamermesh (2018), Angrist et al. (2017) and Angrist et al. (2020). However, it is important to note that these studies used different sample sizes, often smaller, and mostly covered shorter timeframes. Additionally, they employed different categories for fields of research and employed diverse tagging strategies. These trends are observed both in the Top 5 and non-Top 5 general research journals, although there are significant differences in the levels of specialization between the two tiers. Focusing on the Top 5 outlets — whose shares we believed to be more representative of economics as a whole — it can be seen that since 2005, applied research has become the most popular field of economics research, surpassing theory. Figure 1 also suggests that, as of the time of writing this article (mid-2023), applied theory may already be the second most popular category in the Top 5 journals.[25]

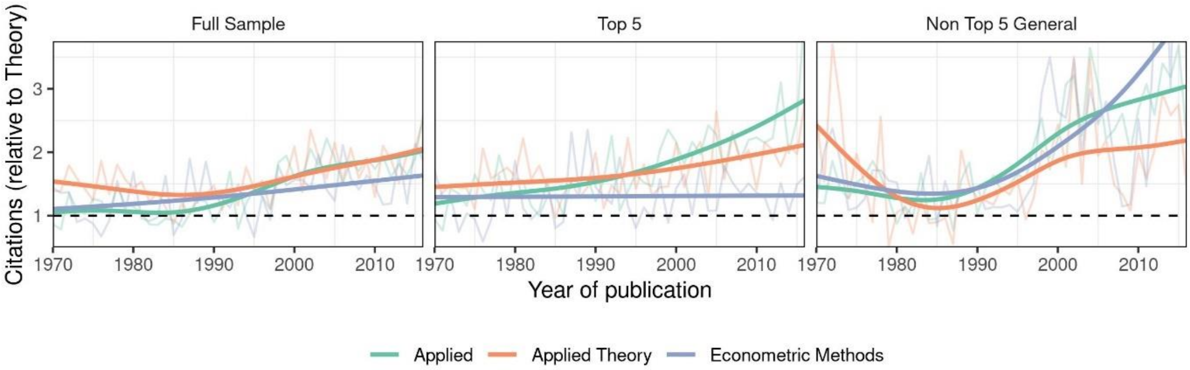
Figure 2 plots trends in received citations for our sample of articles. To obtain this figure, we first calculated the sum of all citations received by articles from field f published in year y. Specifically, we considered citations received within the first seven years since publication.[26] Next, we divided this sum by the number of articles published in year y from field f, resulting in the ratio/average $\bar{c}_y^f$. The plot displays the evolution of the ratios $\bar{c}_y^f / \bar{c}_{theory}^f$ for each year y from 1970 up to and including 2016, encompassing all fields of economics research except theory (which by definition remains at a value of 1 for all years). By dividing by $\bar{c}_{theory}^f$, we control for secular trends in citations, including "citation inflation."[27] Values greater than 1 for a specific field f and year y indicate that, during the first seven years since publication, papers from field f published in year y received more citations on average than theory papers published during the same period.

[25] Recall that Constellate data is available up to 2016.

[26] Analyzing received citations in the first years since publication is a commonly adopted approach in the literature, as it has been observed that citations received during this initial period are highly correlated with citations received over longer timeframes (Hamermesh, 2018; Anauati, Galiani, and Gálvez, 2020).

[27] The tendency of newer articles being cited more than older articles for the very fact of just being newer (see Neff and Olden. 2010 and Galiani and Gálvez, 2019).

Figure 2. Citation counts across fields of economics research relative to theory



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). The dashed horizontal line at 1 serves as a visual guide, indicating the value at which the average citation of a field equals that of theory. We only consider citations received within the first seven years of publication.

Figure 2 shows that theory articles had the lowest citation counts throughout the entire period (all of the plotted curves lie above 1). In the full sample, applied papers experienced a substantial increase in citations compared to theory papers starting from the 1990s.[28] Our data also suggest that applied theory papers were the ones cited the most during the whole period, and that econometric methods articles showed a steady, but milder, growing trend during the whole period. When focusing on the Top 5 outlets, the citation ratio between econometric methods and theory articles remained steady and close to 1. However, applied and applied theory articles displayed a consistently positive slope throughout the period, with a particularly strong trend observed for applied papers. Note that Top 5 applied articles were already the most cited since the mid-1990s. The pattern observed is quite similar for the non-Top 5 general research outlets. However, a growing-in-time gap between citations received by econometric methods articles relative to citations received by theory articles is observed.[29] Given that patterns and trends do not vary between Top 5 and non-Top 5 outlets, from now on, we will only present results for our full sample of articles.

We now study overall trends in the content of articles across different fields of economics research. As explained in Section 3.3, we utilized an LDA model to identify the topics present in our final sample of articles. This analysis allowed us to determine: 1) the topics covered in our corpus, 2) the presence of these topics in each article, and 3) the presence of each word in each topic. With this information, our first objective is to investigate the changes in the diversity of topics addressed within each field of economics research. To accomplish this, we adopt the approach outlined in Hall, Jurafsky, and Manning (2008). Specifically, for each field f and publication year y, we calculate topic entropy as follows:

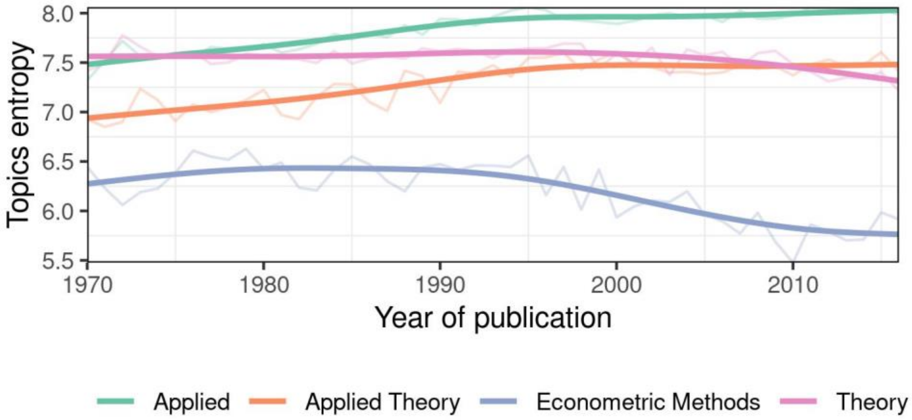
$$H(z|f,y) = -\sum_{k=1}^{K} \hat{p}(z_k|f,y) \cdot \log(\hat{p}(z_k|f,y)),$$

where K represents the total number of topics considered (which in our case is 695), z_k denotes a specific topic, and $\hat{p}(z_k|f,y)$ represents the presence of topic z_k in articles from field f published in year y. To estimate $\hat{p}(z_k|f,y)$, we first compute the ratio between the number of articles from field f published in year y that contain topic z_k and the total number of articles from field f published in year y (denoted as $r_{z_k,f,y}$). Then, we derive $\hat{p}(z_k|f,y)$ as $r_{z_k,f,y}$ divided by the sum of ratios for all topics z_j, where j

[28] The rising gap has been documented before (see, for example, Josh Angrist et al., (2020); Hamermesh, 2018). Although, once again, using different — often smaller — samples of articles and tagging strategies. It is worth noting that this trend aligns with what Angrist and Pischke (2010) refer to as the "credibility revolution in empirical economics," a shift towards enhanced reliability in empirical economics characterized by a strong emphasis on the quality of research design and the use of more experimental and quasi-experimental methods.
[29] This last pattern should be taken with caution, as non-Top 5 general research outlets published very few econometric methods articles since the 2000 (see Figure 1).

ranges from 1 to K (i.e., $\sum_{j=1}^{K} r_{z_j,f,y}$).[30] Recall that we consider an article i as containing a particular topic if that topic is among the top ten topics with the highest presence in the article (see Section 3.3). High values of $H(z|f,y)$ indicate that articles from field f published in year y cover a wide range of topics, while lower values suggest a narrower focus. Figure 3 displays the values of $H(z|f,y)$ across different fields of economics research and publication years.

Figure 3. Topic entropy across fields of economics research



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009).

Figure 3 reveals several noteworthy patterns. First, there are significant variations in topic entropy levels among different fields of economics research, with econometric methods exhibiting the lowest entropy. Second, trends in topic entropy differ across fields. Applied and applied theory articles demonstrated consistent positive trends throughout the entire period, with applied articles exhibiting higher entropy as early as the 1980s. Topic entropy for theory articles remained relatively stable until the mid-1990s, after which a slight decline is observed. Econometric methods articles experienced a modest increase in topic entropy until the early 1990s, followed by a consistent decrease (reaching its lowest point by 2016). Drops in topic entropy are indicative of a field becoming more specialized, while rises are indicative of it becoming more multidisciplinary.

To gain deeper insights into the dynamics depicted in Figure 3 , we delve into the analysis of convergence and divergence in the topics covered by articles from different fields of economics research over time. Note that Figure 3 does not provide a clear indication as to whether applied and applied theory articles are expanding their coverage of topics to the extent that they overlap, or if they cover a wider range of topics with minimal overlap between them. To quantify the level of topic convergence/divergence, we calculated the Jensen-Shannon divergence values across fields of economics research for each year in our sample. Specifically, for each pair of fields f_1 and f_2, and for every publication year y, we calculate the following expression:

$$JSD(\hat{p}(z|f_1,y) \parallel \hat{p}(z|f_2,y)) = \frac{1}{2}D(\hat{p}(z|f_1,y) \parallel M) + \frac{1}{2}D(\hat{p}(z|f_s,y) \parallel M),$$

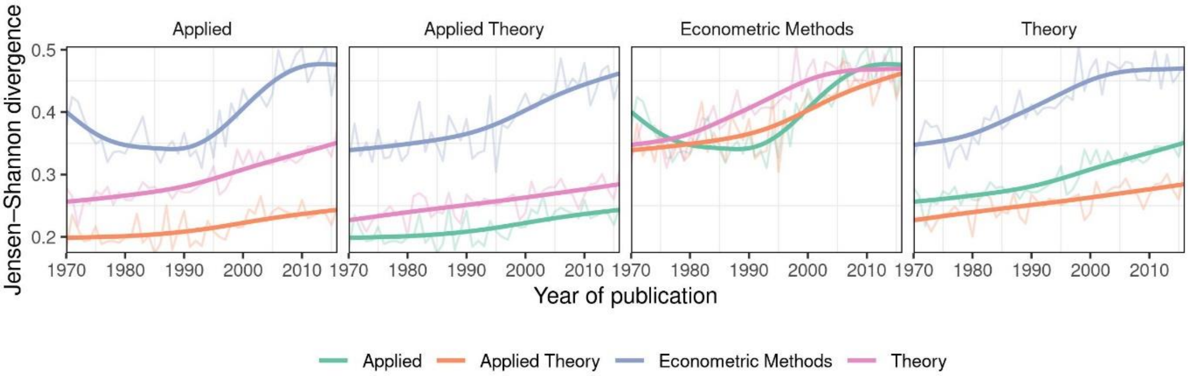
where $\hat{p}(z|f_1,y)$ and $\hat{p}(z|f_2,y)$ are vectors representing topic presence for fields f_1 and f_2, respectively, in year y. These vectors are obtained by concatenating the values of $\hat{p}(z_k|f_1,y)$ and $\hat{p}(z_k|f_2,y)$ across values of k. M is equal to $(\hat{p}(z|f_1,y) + \hat{p}(z|f_2,y))/2$. $D(P \parallel Q)$ represents for the Kullback-Leibler divergence calculated for P with Q as the reference.[31] JSD is a symmetric measure, and high values of

[30] We divide $r_{z_k,f,y}$ by $\sum_{j=1}^{K} r_{z_j,f,y}$ to guarantee that the sum of $\hat{p}(z_k|f,y)$ across all possible values of k equals 1.

[31] The Kullback-Leibler divergence, denoted as $D(P \parallel Q)$, is a measure of the difference between two probability distributions. it quantifies how distribution P differs from distribution Q. It is non-symmetric and undefined when there exists an index k for which $P_k > 0$ and $Q_k = 0$. The two limitations are overcome by the Jensen-Shannon divergence, which is a symmetrized version of the Kullback-Leibler divergence.

JSD indicate a greater dissimilarity or divergence between the probability distributions being compared (in our case, that the fields have distinct topic profiles). In Figure 4 we present the trends in JSD for all pairs of fields and for all years considered in this study. Each panel in the figure corresponds to a field of economics research, and within each panel, the JSD is calculated for every other field of economics research and every year considered in the study.

Figure 4. Jensen-Shannon divergence across fields of economics research

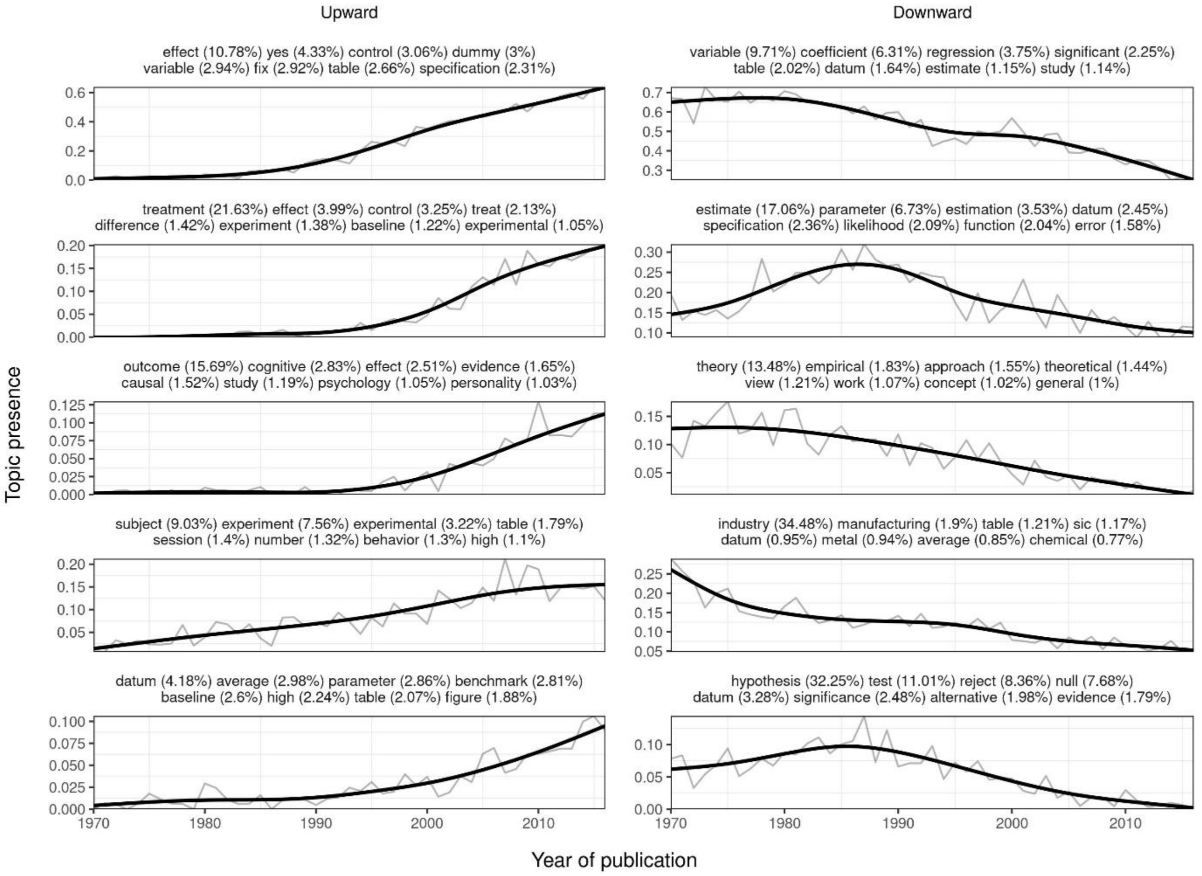


Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009).

Figure 4 depicts that, irrespective of the field under consideration, the JSD between each field and any other field of economics research exhibit a consistent positive trend. This indicates that over time, all fields of economics research tended to specialize in different topics. Looking at the specific JSD levels, we observe that the field of econometric methods showed the highest topic divergence compared to any other field. On the other hand, the field of applied theory demonstrated closer topic profiles to both applied and theory fields. This finding suggests that applied theory stands between the applied and theory fields in terms of topic similarity, occupying an intermediate position. These patterns, along with the observations from Figure 3, further suggest that both theory and econometric methods articles are becoming more specialized. In the case of applied and applied theory, the patterns suggest that while both fields are covering a wider set of topics, there is no tendency for them to overlap.

We now study which topics gained and lost prominence across fields of economics research since the mid-1990s. We identify variation in topic presence for a given field of economics research in the following way. First, for each field of economics research f and every detected topic z_k, we fit a linear regression model to the evolution of $\hat{p}(z_k|f, y)$ over y where y ranges from 1995 up to and including 2016. This results in a set of regression slopes, where positive values indicate an increasing trend in topic presence and negative values indicate a decreasing trend. In Figure 5, the leftmost panels depict the evolution of the five topics with the highest positive slopes, which represent a rising prominence in our sample of applied papers since 1995. Conversely, the rightmost panels display the evolution of the five topics with the lowest (negative) slopes.

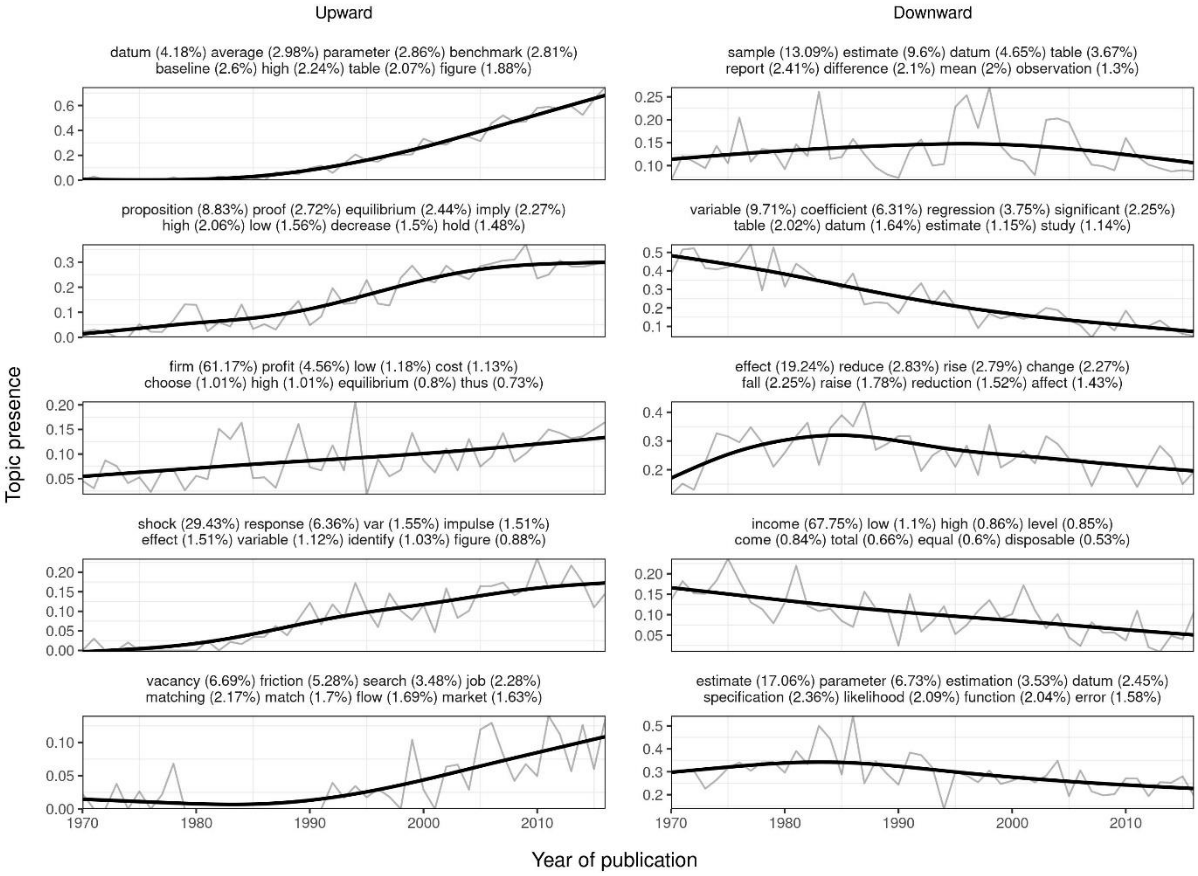
Figure 5. Topics gaining and losing more presence in applied papers since 1995



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). Each panel title displays the top 8 words that make up the corresponding topic, along with their participation within the topic (in parentheses). In the left/right column of panels, higher panels represent topics with higher/lower estimated variations (i.e., larger/lower slope values).

Figure 5 illustrates a strong shift in the contents of applied papers. The topics showing the highest rising trends are closely related to impact analysis and causal analysis, including words such as *effect*, *treatment*, and *control*. By 2016, over 60% of all applied articles included the topic dominated by the word *effect*. On the other hand, topics associated with simple correlational analysis, indicated by words like *variable*, *coefficient*, *regression*, *table*, *specification*, and *significant*, experienced a significant decrease in presence. Figure 5 also indicates a shift not only in statistical techniques and empirical strategies but also in domains of study. The presence of a topic related to experimental economics, including words like *subject*, *experiment*, and *experimental*, saw a substantial rise, while a topic encompassing words like *industry* and *manufacture* showed a marked decline.

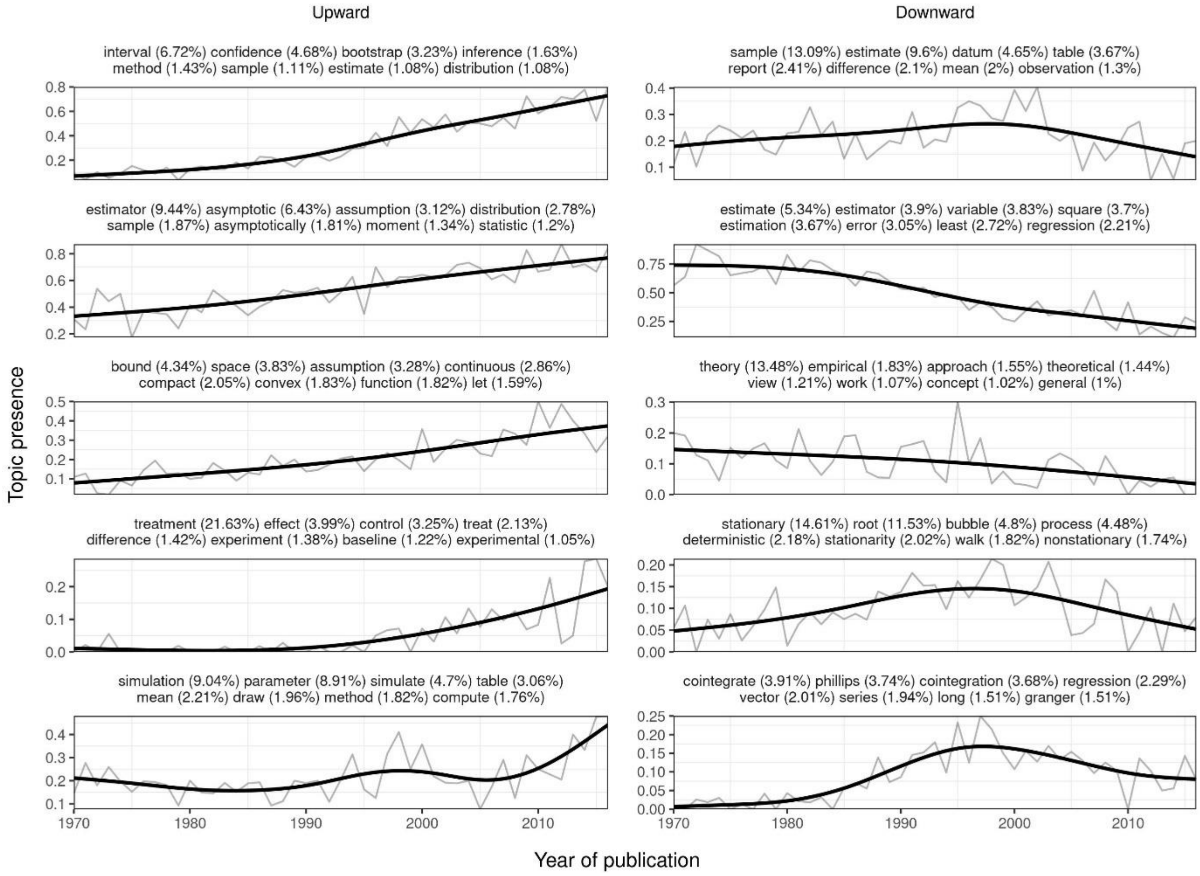
Figure 6. Topics gaining and losing more presence in applied theory papers since 1995



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). Each panel title displays the top 8 words that make up the corresponding topic, along with their participation within the topic (in parentheses). In the left/right column of panels, higher panels represent topics with higher/lower estimated variations (i.e., larger/lower slope values).

Figure 6 replicates Figure 5, focusing on applied theory articles. The results show that topics commonly associated with modern macroeconomics research have significantly increased their presence since 1995. These topics include words such as *shock*, *response*, *impulse*, *friction*, and *search*. Additionally, the topic centered around the word *firm* also experienced a notable rise during the same period. Interestingly, two other topics that gained prominence are 1) a topic related to data and estimation (involving words like *datum*, *average*, and *parameter*) and 2) a topic associated with theory and demonstrations (including words such as *proposition*, *proof*, and *equilibrium*). These two topics reflect the middle-ground nature of this field, combining theoretical and applied analyses. Similar to applied articles, several topics associated with simple correlational analysis showed a decline in their presence during the period.

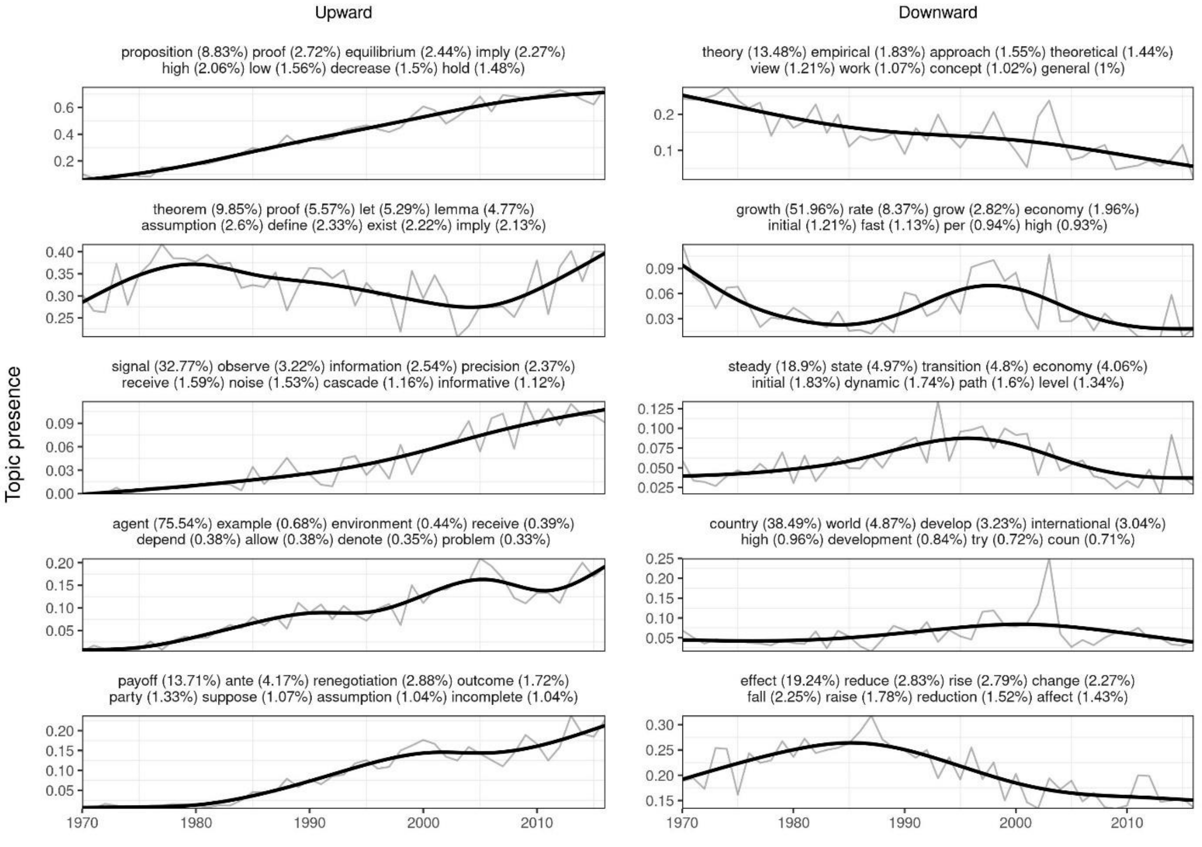
Figure 7. Topics gaining and losing more presence in econometric methods papers since 1995



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). Each panel title displays the top 8 words that make up the corresponding topic, along with their participation within the topic (in parentheses). In the left/right column of panels, higher panels represent topics with higher/lower estimated variations (i.e., larger/lower slope values).

Figure 7 displays the topics that have gained and lost prominence in econometric methods articles. Notably, two topics related to computational statistics, involving words like *bootstrap* and *simulation*, have experienced the most significant increase since 1995. Moreover, there has been a drastic positive trend in a topic that includes words such as *estimator* and *asymptotic*, with approximately 80% of all econometric methods articles addressing this topic by 2016 (which could easily be associated to asymptotic properties of estimators). Additionally, around 50% of econometrics articles cover a topic that encompasses the words *bound*, *space*, and *assumption*, which are commonly used jargon in the study of estimators bounds. Interestingly, albeit to a lesser extent, a topic strongly associated with causal analysis (including words like *treatment*, *effect*, and *control*) has also increased over the period. Identifying clear patterns in the topics that have lost prominence since 1995 is more challenging. However, it is worth noting that two of these topics are closely related to time series analysis, with one involving words like *stationary* and *root* and the other including words such as *cointegrate*, *series*, and *granger*.

Figure 8. Topics gaining and losing more presence in theory methods papers since 1995



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). Each panel title displays the top 8 words that make up the corresponding topic, along with their participation within the topic (in parentheses). In the left/right column of panels, higher panels represent topics with higher/lower estimated variations (i.e., larger/lower slope values).

Finally, Figure 8 highlights the topics that have gained and lost prominence since 1995 in theory articles. The two topics that have experienced the most significant increase are closely related to formal mathematical proofs, with one encompassing words like *proposition*, *proof*, and *equilibrium*, and the other including words such as *proof*, *lemma*, and *assumption*. Two additional topics are associated with game theory studies, with one featuring words like *signal*, *observe*, and *information*, and the other including the word *payoff*. Another topic that primarily uses the word *agent* has also seen a rise since 1995. Interestingly, the topic that has experienced the sharpest decrease is predominantly composed of the word *theory*.[32] Other topics showing declining trends include a topic associated with steady-state analysis (including words like *steady*, *state*, and *path*), a topic potentially linked to growth studies (encompassing words like *growth*, *rate*, and *grow*), and a topic mainly consisting of the word *effect*.

4.2. The interplay between fields of economics research and fields of study other than economics

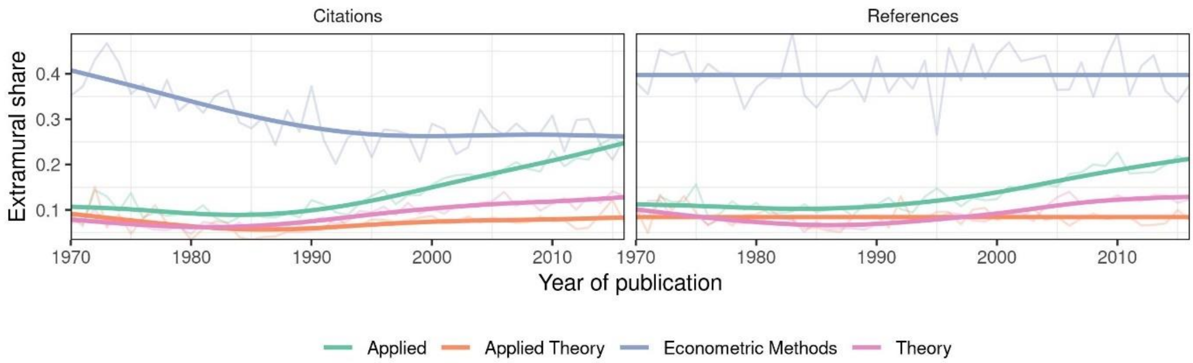
Up to now, we focused on studying aggregated trends across fields of economics research. In this section, we shift our focus to exploring the interactions between economics research articles and articles from other fields of study, such as medicine and psychology. Previous studies that have explored extramural citations include Pieters and Baumgartner (2002), Fourcade, Ollion, and Algan (2015), and Angrist et al. (2020). Among these studies, Angrist et al. (2020) investigated differences in extramural citations across different fields within economics research. Here, we extend this analysis by

[32] Upon manual inspection, we observed that many of these papers emphasize the field of economics itself and use the term "economic theory" to refer to the body of knowledge generated by economics, rather than referring to the approach of studying phenomena using a theory-based strategy.

focusing on documenting trends consistent with specialization within economics and incorporating topic analysis.

In Figure 9, we present two panels. The leftmost panel illustrates the trends in the proportion of extramural citations received by the average article within each field of economics research. The rightmost panel shows the trends in the proportion of extramural references (citations to articles from other fields generated by articles in our sample) across different fields of economics research.[33]

Figure 9. Share of extramural citations and references across fields of economics research



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). We only consider citations received within the first seven years of publication.

Figure 9 illustrates the variability in the extent to which economics articles are cited by and reference extramural articles across fields of economics research. Econometric methods papers receive a higher share of extramural citations and reference more extramural articles than any other field. However, while the share of extramural references has remained remarkably stable throughout the entire period, the share of extramural citations has experienced a steady decline. Applied is the field that comes second in terms of extramural citations and references. For applied papers, we observe a steady increase in both extramural citations and extramural references since the early 1990s. By 2016, approximately 25% of all citations received by applied papers are from outside the field, and nearly 20% of their references are extramural ones. Finally, both theory and applied theory papers display a more stable behavior during the period 1970-2016, with theory papers showing a mild but negligible upward trend and applied theory papers showing no tendency at all.

To better understand these patterns, Table 6 presents the top twenty topics that exhibit the strongest correlation with the articles' shares of extramural citations (additionally, it provides the percentage of articles containing each listed topic for each field of economics research). The calculation of these correlations is as follows. First, we determine the share of extramural citations received within the first seven years of publication for each article in our complete sample. Next, we calculate the Pearson correlation between the articles' shares of extramural citations and the presence of each topic across the different articles (yielding one correlation for each topic considered). Finally, we identify the topics with the highest correlation values, indicating their strong association with extramural citations.

[33] Concretely, we calculate the percentage of citations/references in our dataset that are attributed to articles from fields outside of economics. When calculating the proportion of extramural citations/references, we exclude articles that have not received at least one citation/reference. For citations, we only consider those received within the first seven years since publication. Subsequently, we calculate the average of these ratios for each field of economics research and year of publication, and plot them accordingly. A citation/reference is categorized as extramural if the Semantic Scholar fields of study tags associated with the citing/referenced article do not include economics.

Table 6. Topics exhibiting the highest correlation with extramural citations

Topic	Correlation with the share of extramural citations	Share of articles including the topic			
		Applied	Applied Theory	Econometric Methods	Theory
outcome (15.69%) cognitive (2.83%) effect (2.51%) evidence (1.65%) causal (1.52%) study (1.19%) psychology (1.05%) personality (1.03%)	0.244	1.72%	0.20%	1.22%	1.15%
score (23.13%) grade (10.71%) test (4.2%) math (1.94%) high (1.68%) standard (1.41%) ged (1.13%) exam (1.07%)	0.226	2.47%	0.15%	1.12%	0.22%
teacher (12.92%) student (4.79%) achievement (4.29%) school (2.43%) teach (2.36%) effect (2.01%) classroom (2%) pupil (1.96%)	0.208	1.87%	0.15%	0.20%	0.29%
school (31.3%) student (4.62%) high (3.43%) education (1.67%) attend (1.66%) enrollment (1.37%) effect (1.34%) dropout (1.06%)	0.205	4.02%	0.95%	0.20%	0.41%
estimator (9.44%) asymptotic (6.43%) assumption (3.12%) distribution (2.78%) sample (1.87%) asymptotically (1.81%) moment (1.34%) statistic (1.2%)	0.202	0.22%	0.65%	55.51%	0.34%
health (24.42%) illness (1.77%) medical (1.58%) mental (1.11%) age (0.96%) disease (0.85%) effect (0.84%) status (0.79%)	0.185	2.98%	1.15%	0.10%	0.26%
effect (10.78%) yes (4.33%) control (3.06%) dummy (3%) variable (2.94%) fix (2.92%) table (2.66%) specification (2.31%)	0.184	28.28%	3.49%	1.94%	0.14%
student (28.71%) class (27.59%) course (6.05%) instructor (0.89%) stu (0.66%) dent (0.64%) section (0.57%) semester (0.53%)	0.179	1.42%	0.35%	0.00%	0.67%
crime (16.54%) criminal (4.03%) police (3.8%) arrest (3.19%) rate (2.31%) offender (1.56%) violent (1.53%) property (1.39%)	0.175	1.82%	0.40%	0.00%	0.65%
patient (22.35%) medical (9.26%) physician (8.59%) doctor (6.46%) practice (5.05%) medicine (1.62%) visit (1.38%) clinical (1.2%)	0.155	0.79%	0.20%	0.10%	0.24%
treatment (21.63%) effect (3.99%) control (3.25%) treat (2.13%) difference (1.42%) experiment (1.38%) baseline (1.22%) experimental (1.05%)	0.137	7.23%	0.65%	5.41%	0.31%
sample (13.09%) estimate (9.6%) datum (4.65%) table (3.67%) report (2.41%) difference (2.1%) mean (2%) observation (1.3%)	0.136	19.56%	5.24%	10.41%	0.19%
black (25.93%) white (22.33%) race (10.33%) racial (3.83%) difference (1.7%) table (0.72%) gap (0.72%) among (0.56%)	0.135	2.04%	0.40%	0.10%	0.22%
hospital (22.71%) patient (3.22%) admission (2.82%) medicare (2.3%) hmo (1.65%) bed (1.63%) health (1.6%) roo (1.04%)	0.135	1.04%	0.15%	0.10%	0.22%
social (17.34%) peer (5.15%) interaction (4.97%) effect (4.85%) friend (1.75%) influence (1.71%) group (0.89%) behavior (0.87%)	0.132	1.93%	1.10%	0.41%	1.56%
inquiry (30.25%) phone (5.85%) association (4.19%) fax (3.99%) turnover (2.7%) find (2.46%) western (2.26%) tio (2.14%)	0.131	0.13%	0.00%	0.00%	0.02%
survey (17.94%) response (8.91%) respondent (8.35%) question (8.18%) ask (4.53%) answer (3.01%) report (2.17%) wtp (1.33%)	0.130	2.74%	0.20%	1.02%	0.19%
performance (26.22%) manager (22.74%) management (7.71%) managerial (4.95%) practice (1.56%) manage (1.55%) perform (1.45%) mance (1.23%)	0.129	0.78%	0.45%	0.20%	1.27%
nonparametric (5.3%) estimation (5.15%) kernel (4.82%) function (4.3%) moment (4.06%) parametric (3.6%) semiparametric (2.81%) estimator (2.34%)	0.129	0.22%	0.75%	21.43%	0.19%
college (26.43%) high (5.2%) student (3.77%) graduate (3.63%) admission (3.24%) tuition (2.39%) enrollment (1.79%) attend (1.75%)	0.127	1.16%	0.90%	0.10%	0.26%

Notes. Topics are sorted in decreasing order according to the estimated correlations. We only consider citations received within the first seven years of publication.

Table 6 reveals that many of the topics strongly associated with extramural citations consist of words commonly linked to fields of study such as psychology, health, criminology, and education. Interestingly, the majority of these topics are found most frequently in applied papers, with the exception of those topics specifically related to econometric studies (containing words such as *estimator* and *nonparametric*), which unsurprisingly are predominantly present in econometric methods papers. Overall, the table suggests that applied articles attract citations from a diverse range of fields. Notably, neither theory nor applied theory papers exhibit the highest presence in any of the topics listed in Table 6.

Table 7 replicates Table 6, but presents the twenty topics that exhibit the strongest negative correlation with the articles' share of extramural citations.

Table 7. Topics exhibiting the lowest correlation with extramural citations

Topic	Correlation with extramural citations	Share of articles including the topic			
		Applied	Applied Theory	Econometric Methods	Theory
equilibrium (43.53%) exist (1.87%) existence (1.09%) unique (0.97%) equilib (0.94%) rium (0.87%) equi (0.82%) librium (0.68%)	-0.207	0.62%	2.70%	0.41%	17.18%
consumption (54.89%) intertemporal (3.64%) consume (3.06%) sumption (2.08%) consump (1.98%) con (1.82%) tion (0.8%) marginal (0.7%)	-0.183	1.84%	4.84%	0.10%	1.58%
real (49.69%) nominal (13.39%) rate (1.62%) interest (1.43%) level (0.72%) price (0.7%) anticipate (0.7%) effect (0.69%)	-0.171	0.97%	1.80%	0.31%	0.67%
output (70.83%) level (3.02%) produce (2.3%) total (1.15%) function (0.95%) put (0.89%) assumption (0.7%) thus (0.54%)	-0.163	0.51%	0.60%	0.51%	0.22%
aggregate (44.95%) tfp (5.03%) across (1.56%) aggregation (1.31%) aggre (1.19%) gate (1.14%) decomposition (1.06%) average (0.96%)	-0.161	0.68%	1.40%	0.51%	0.14%
capital (63.92%) physical (2.19%) accumulation (2.14%) return (1.94%) stock (1.06%) investment (0.97%) economy (0.5%) neoclassical (0.49%)	-0.158	1.44%	4.94%	0.31%	2.23%
economy (28.96%) pareto (7.79%) competitive (2.89%) feasible (1.55%) initial (1.36%) endowment (1.29%) every (1.27%) equilibrium (1.18%)	-0.156	0.21%	0.50%	0.00%	3.50%
datum (4.18%) average (2.98%) parameter (2.86%) benchmark (2.81%) baseline (2.6%) high (2.24%) table (2.07%) figure (1.88%)	-0.155	1.66%	27.46%	0.61%	0.48%
shock (29.43%) response (6.36%) var (1.55%) impulse (1.51%) effect (1.51%) variable (1.12%) identify (1.03%) figure (0.88%)	-0.154	3.01%	8.84%	3.27%	0.72%
proposition (8.83%) proof (2.72%) equilibrium (2.44%) imply (2.27%) high (2.06%) low (1.56%) decrease (1.5%) hold (1.48%)	-0.147	0.51%	14.38%	3.16%	40.45%
rate (64.44%) interest (11.57%) high (1.03%) expect (0.88%) level (0.79%) low (0.63%) change (0.5%) constant (0.42%)	-0.147	2.74%	5.09%	0.51%	2.11%
growth (51.96%) rate (8.37%) grow (2.82%) economy (1.96%) initial (1.21%) fast (1.13%) per (0.94%) high (0.93%)	-0.144	3.58%	4.84%	0.31%	2.66%
inflation (39.81%) monetary (2.45%) inflationary (1.87%) high (1.6%) phillips (1.55%) policy (1.39%) nominal (1.23%) policymaker (1.16%)	-0.141	1.74%	4.94%	0.61%	1.99%
monetary (62.27%) multiplier (10.72%) authority (2.01%) aggregate (1.46%) mone (1.08%) simple (1.05%) tary (1.02%) interest (0.81%)	-0.140	0.21%	0.45%	0.00%	0.36%
price (69.94%) change (1.36%) average (0.52%) low (0.48%) high (0.38%) different (0.33%) cost (0.33%) demand (0.28%)	-0.139	7.20%	8.34%	1.94%	7.05%
trade (63.67%) international (1.65%) bilateral (1.54%) trading (1.08%) world (1.08%) flow (0.75%) barrier (0.75%) good (0.75%)	-0.138	1.74%	3.44%	0.10%	2.47%
steady (18.9%) state (4.97%) transition (4.8%) economy (4.06%) initial (1.83%) dynamic (1.74%) path (1.6%) level (1.34%)	-0.137	0.27%	4.49%	0.31%	4.10%
effect (19.24%) reduce (2.83%) rise (2.79%) change (2.27%) fall (2.25%) raise (1.78%) reduction (1.52%) affect (1.43%)	-0.137	2.94%	10.78%	1.33%	9.25%
asset (64.48%) portfolio (7.8%) hold (5.14%) holding (4.32%) riskless (0.87%) financial (0.69%) total (0.66%) composition (0.43%)	-0.131	0.46%	1.20%	0.00%	1.22%
labor (57.82%) supply (7.62%) force (4.97%) wage (1.86%) work (1.56%) rate (0.84%) income (0.65%) market (0.63%)	-0.129	1.03%	3.20%	0.10%	0.53%

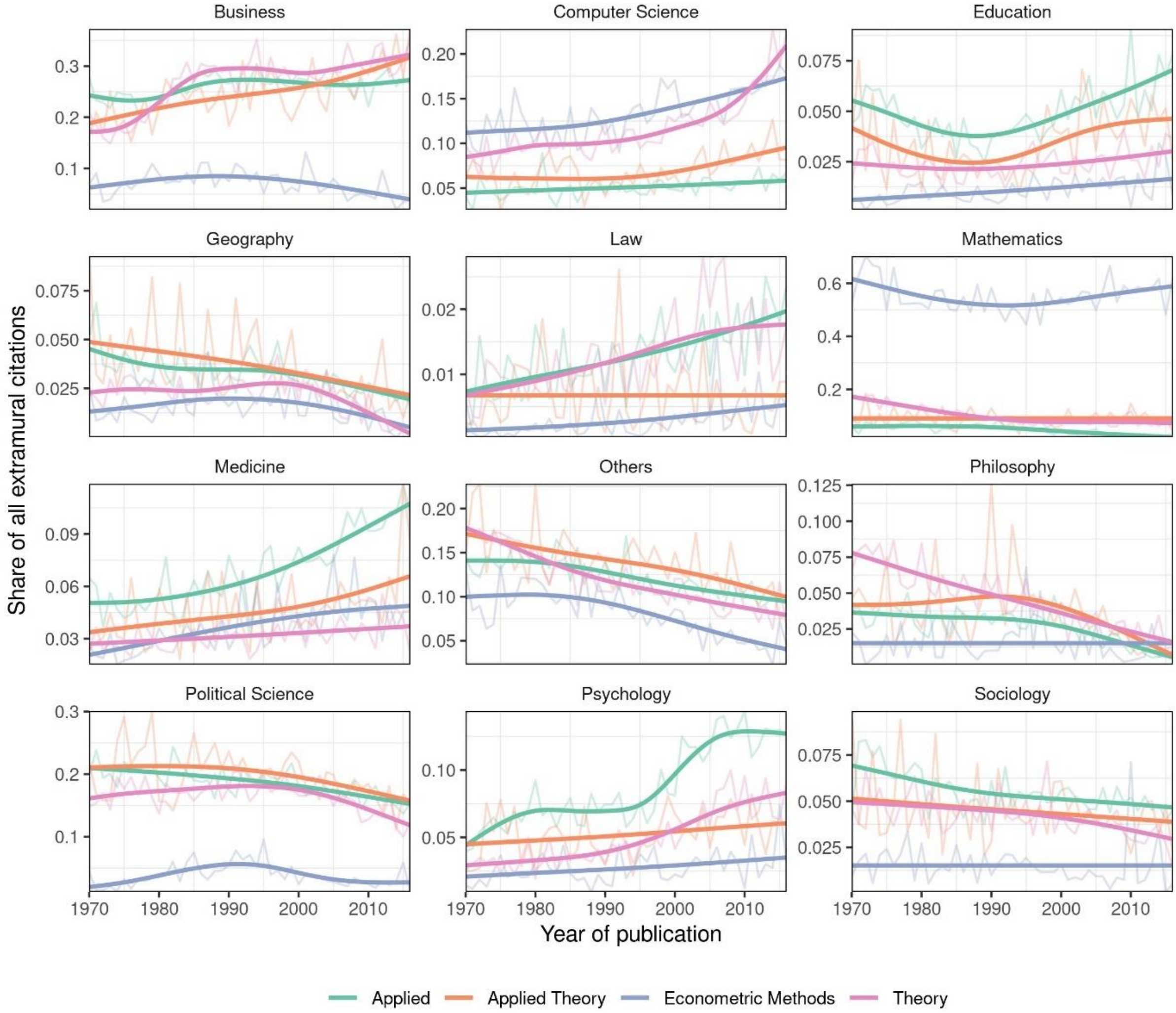
Notes. Topics are sorted in increasing order according to the estimated correlations. We only consider citations received within the first seven years of publication.

Table 7 illustrates that the topics with lower correlations to extramural citations predominantly consist of words that can be categorized as economics jargon. These topics are primarily found in applied theory and theory papers. Notably, there are no instances where applied or econometric methods papers exhibit the highest topic presence among the listed topics.

To further explore the interaction between economics and other fields of study, Figure 10 visualizes how different fields of study interact with various fields of economics research. This analysis was conducted as follows. First, for each article i in our sample that received at least one extramural citation (within its first seven years since publication) and for each field of study, we calculated the proportion of i's extramural citations received within the first seven years since publication that originate from each field of study. These proportions represent the share of extramural citations attributed to each field of study (e.g., around 20% of i's extramural citations come from articles in the field of medicine, around 15% from psychology articles, and so on).[34] Then, having calculated these shares, we computed the average for each year of publication and for each field of economics research. Figure 10 displays these averages across fields of study, fields of economics research, and years of publication.

[34] Note that an extramural article j citing an article i can be assigned multiple fields of study tags by Semantic Scholar (e.g., medicine and psychology). To address this, when calculating the averages mentioned earlier, we consider that extramural articles belonging to k different fields of study contribute a fraction of $1/k$ to the numerator of the share/average calculation.

Figure 10. Share of extramural citations received from non-economics fields by year of publication and field of economics research



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). “Other” includes: Agricultural and food sciences, art, biology, chemistry, engineering, environmental science, geology, history, linguistics, materials science, and physics. We only consider citations received within the first seven years of publication.

Figure 10 shows several interesting patterns in the origin of extramural citations to economics papers. First, the origin of citations varies greatly across fields of economics research. For example, mathematics represents a large share of econometric methods citations. Computer science represents a large share of theory and econometric methods citations (a pattern also documented in Angrist et al., 2020). Business is the field of study that cites applied, applied theory, and theory the most, almost not citing econometric methods at all. Second, and perhaps more interesting, trends in the origin of citations vary greatly across fields of economics research and fields of study. For example, there has been a large increase in the importance of citations coming from medicine for applied papers. By 2016, medicine represented more than 10% of all applied papers' extramural citations. A similar pattern is observed for citations coming from psychology, law, and to a lesser extent, education. For theory, strong rising trends can be observed for citations coming from computer science and law. On the other hand. some fields of study seem to have lost importance in terms of citations to economics papers. This is the case for political science, philosophy, geography, sociology, and the disciplines included in "Others" (when taken as a whole).

To gain further insights into the patterns depicted in Figure 10, we present Table 8, which showcases the topics exhibiting the strongest correlation with extramural citations from each field of study. The methodology employed in constructing this table is akin to that used for Table 6, with the difference that it considers the share of extramural citations from a particular field of study, rather than the share of received extramural citations relative to all received citations. Additionally, Table 8 provides the proportion of applied papers featuring each topic.

Table 8. Topics associated with citations from various fields of study

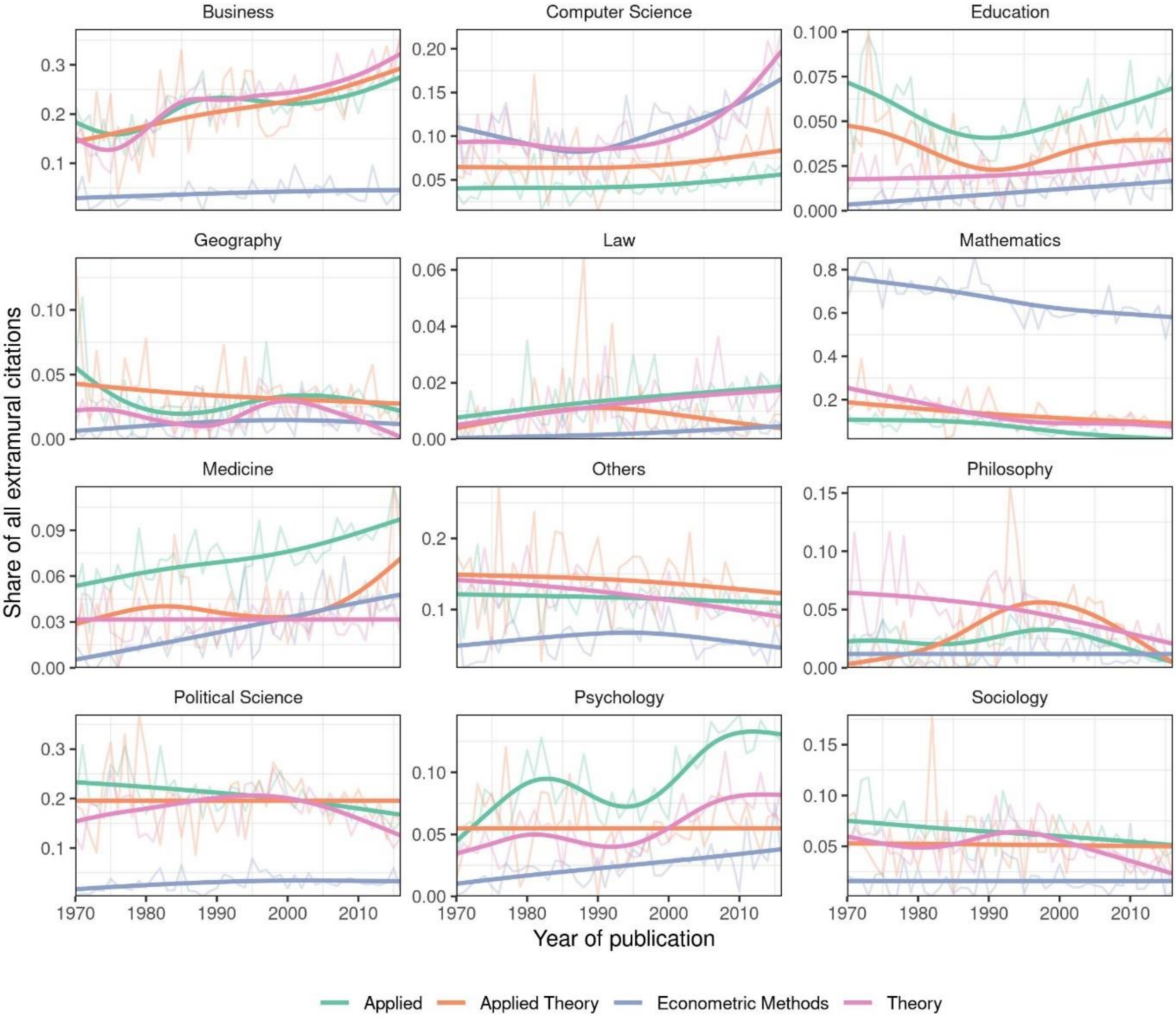
Field of study	Topic	Correlation with citations shares	Share of articles including the topic			
			Applied	Applied Theory	Econometric Methods	Theory
Business	firm (36.41%) datum (1.9%) average (1.1%) sample (1.07%) table (0.99%) number (0.86%) measure (0.85%) total (0.81%)	0.335	7.74%	4.96%	0.39%	0.38%
	performance (26.22%) manager (22.74%) management (7.71%) managerial (4.95%) practice (1.56%) manage (1.55%) perform (1.45%) mance (1.23%)	0.231	1.75%	0.90%	0.34%	1.62%
	shareholder (7.41%) takeover (2.8%) governance (2.11%) control (1.79%) tender (1.63%) journal (1.28%) corporate (1.17%) shleifer (1.03%)	0.227	1.10%	0.73%	0.05%	1.85%
Computer Science	game (17.82%) player (15.3%) payoff (5.12%) equilibrium (4.65%) play (4.58%) strategy (1.63%) outcome (1.23%) subgame (0.93%)	0.239	3.02%	1.80%	0.54%	9.92%
	step (22.89%) algorithm (15.92%) number (3.76%) shapley (3.51%) grid (2.65%) visit (2.4%) compute (2.16%) find (2.01%)	0.235	0.26%	0.31%	3.45%	1.31%
	theorem (9.85%) proof (5.57%) let (5.29%) lemma (4.77%) assumption (2.6%) define (2.33%) exist (2.22%) imply (2.13%)	0.195	0.39%	4.16%	43.62%	33.49%
Education	school (31.3%) student (4.62%) high (3.43%) education (1.67%) attend (1.66%) enrollment (1.37%) effect (1.34%) dropout (1.06%)	0.503	4.62%	1.46%	0.25%	0.48%
	teacher (12.92%) student (4.79%) achievement (4.29%) school (2.43%) teach (2.36%) effect (2.01%) classroom (2%) pupil (1.96%)	0.463	2.22%	0.42%	0.25%	0.27%
	college (26.43%) high (5.2%) student (3.77%) graduate (3.63%) admission (3.24%) tuition (2.39%) enrollment (1.79%) attend (1.75%)	0.431	1.88%	1.04%	0.10%	0.34%
Geography	city (46.55%) urban (3.87%) population (1.45%) smsa (1.11%) live (0.9%) york (0.86%) metropolitan (0.82%) across (0.76%)	0.219	2.15%	1.42%	0.00%	0.41%
	location (23.88%) spatial (8.68%) locate (5.42%) center (2.99%) agglomeration (2.93%) distance (2.15%) commute (1.81%) geographic (1.46%)	0.178	1.42%	1.42%	0.79%	1.26%
	migration (29.62%) migrant (13.25%) origin (5.58%) migrate (4.68%) destination (4.2%) emigration (2.08%) united (1.35%) flow (1.21%)	0.151	1.31%	0.31%	0.05%	0.33%
Law	crime (16.54%) criminal (4.03%) police (3.8%) arrest (3.19%) rate (2.31%) offender (1.56%) violent (1.53%) property (1.39%)	0.588	1.77%	0.52%	0.25%	0.50%
	court (10.55%) judge (8.34%) sentence (4%) defendant (2.88%) judicial (2.5%) justice (2.28%) incarceration (2.01%) convict (1.92%)	0.458	0.81%	0.14%	0.10%	0.58%
	etal (13.38%) settlement (11.93%) award (7.18%) dispute (4.04%) plaintiff (3.82%) litigation (3.64%) settle (2.95%) suit (2.11%)	0.217	0.52%	0.28%	0.05%	0.40%
Mathematics	estimator (9.44%) asymptotic (6.43%) assumption (3.12%) distribution (2.78%) sample (1.87%) asymptotically (1.81%) moment (1.34%) statistic (1.2%)	0.547	0.55%	1.46%	51.06%	0.41%
	matrix (25.25%) vector (19.12%) element (7.43%) diagonal (3.13%) covariance (3%) coefficient (1.5%) structure (1.35%) zero (1.35%)	0.424	0.65%	2.08%	36.27%	3.20%
	estimate (5.34%) estimator (3.9%) variable (3.83%) square (3.7%) estimation (3.67%) error (3.05%) least (2.72%) regression (2.21%)	0.422	7.43%	3.57%	54.26%	0.64%
Medicine	health (24.42%) illness (1.77%) medical (1.58%) mental (1.11%) age (0.96%) disease (0.85%) effect (0.84%) status (0.79%)	0.476	4.04%	1.21%	0.25%	0.47%
	patient (22.35%) medical (9.26%) physician (8.59%) doctor (6.46%) practice (5.05%) medicine (1.62%) visit (1.38%) clinical (1.2%)	0.394	1.03%	0.35%	0.10%	0.29%
	hospital (22.71%) patient (3.22%) admission (2.82%) medicare (2.3%) hmo (1.65%) bed (1.63%) health (1.6%) roo (1.04%)	0.387	1.22%	0.31%	0.20%	0.21%
Others	century (5.42%) historical (3.82%) history (3.41%) press (2.26%) early (1.83%) nineteenth (1.73%) modern (1.49%) cambridge (1.2%)	0.062	0.32%	0.13%	0.00%	0.28%
	pollution (11.78%) environmental (11%) emission (9.1%) air (6.22%) abatement (2.42%) permit (2.2%) clean (2.14%) level (1.92%)	0.055	0.33%	0.41%	0.04%	0.24%
	electricity (8.22%) power (5.42%) electric (4.4%) gas (2.57%) energy (1.81%) generator (1.8%) load (1.71%) generation (1.61%)	0.051	0.48%	0.19%	0.09%	0.06%
Philosophy	theory (13.48%) empirical (1.83%) approach (1.55%) theoretical (1.44%) view (1.21%) work (1.07%) concept (1.02%) general (1%)	0.120	7.25%	6.38%	10.94%	14.94%
	religious (11.57%) religion (8.12%) church (7.79%) catholic (6.41%) protestant (3.41%) sacrifice (2.33%) catholics (1.48%) secular (1.45%)	0.062	0.61%	0.24%	0.05%	0.11%
	welfare (33.92%) social (15.68%) alternative (1.75%) pareto (1.31%) socially (1.19%) function (0.92%) choice (0.85%) sen (0.66%)	0.059	1.15%	4.72%	1.18%	7.24%
Political Science	political (15.51%) power (6.11%) politician (3.69%) citizen (3.05%) public (1.75%) politic (1.5%) democracy (0.96%) persson (0.9%)	0.303	2.34%	2.39%	0.15%	3.62%
	vote (26.3%) voter (13.19%) outcome (1.55%) turnout (1.42%) election (1.34%) issue (1.06%) number (0.97%) pivotal (0.94%)	0.246	2.06%	2.12%	0.10%	3.19%
	election (17.22%) party (6.31%) electoral (6.26%) political (3.89%) incumbent (2.57%) presidential (2.56%) partisan (1.57%) reelection (1.38%)	0.222	1.60%	1.25%	0.10%	0.64%
Psychology	outcome (15.69%) cognitive (2.83%) effect (2.51%) evidence (1.65%) causal (1.52%) study (1.19%) psychology (1.05%) personality (1.03%)	0.375	3.53%	0.66%	1.58%	1.45%
	mother (10.63%) child (10.3%) effect (2.67%) father (1.39%) maternal (1.35%) age (1.18%) sample (0.99%) time (0.85%)	0.291	2.58%	1.01%	0.25%	0.13%
	subject (9.03%) experiment (7.56%) experimental (3.22%) table (1.79%) session (1.4%) number (1.32%) behavior (1.3%) high (1.1%)	0.290	9.41%	1.98%	3.84%	1.37%
Sociology	woman (33.63%) man (22.15%) work (1.81%) marry (1.64%) force (1.42%) difference (1.3%) less (1%) single (0.93%)	0.182	4.07%	1.73%	0.44%	0.45%
	religious (11.57%) religion (8.12%) church (7.79%) catholic (6.41%) protestant (3.41%) sacrifice (2.33%) catholics (1.48%) secular (1.45%)	0.169	0.61%	0.24%	0.05%	0.11%
	crime (16.54%) criminal (4.03%) police (3.8%) arrest (3.19%) rate (2.31%) offender (1.56%) violent (1.53%) property (1.39%)	0.166	1.77%	0.52%	0.25%	0.50%

Notes. "Other" includes: Agricultural and food sciences, art, biology, chemistry, engineering, environmental science, geology, history, linguistics, materials science, and physics. Within each citing field of study, topics are sorted in decreasing order according to the estimated correlations. We only consider citations received within the first seven years of publication.

Table 8 shows that different fields of study tend to cite articles containing different topics. The patterns observed are the ones that one would expect beforehand. For instance, education articles tend to cite economics articles featuring topics containing the words *school*, *teacher*, and *college*. Similarly, medicine articles tend to cite economics articles encompassing topics containing words such as *health*, *patient*, and *hospital*, while computer science articles cite economics articles with topics centered around words such as *game*, *algorithm*, and *theorem*. Figure 10 and Table 8 exhibit consistent patterns. Within a field of study that predominantly cites a given field of economics research (e.g., medicine citing applied papers), the topics cited the most are predominantly present in articles from the same field of economics research (e.g., the topic led by the word "health" is present the most in applied papers).

Finally, Figure 11 examines the citation behavior of different fields of economics research towards other fields of study. It presents a similar analysis to Figure 10, but focuses on the share of extramural references to specific fields of study.

Figure 11. Share of extramural references to non-economics fields by year of publication and field of economics research



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). “Other” includes: Agricultural and food sciences, art, biology, chemistry, engineering, environmental science, geology, history, linguistics, materials science, and physics.

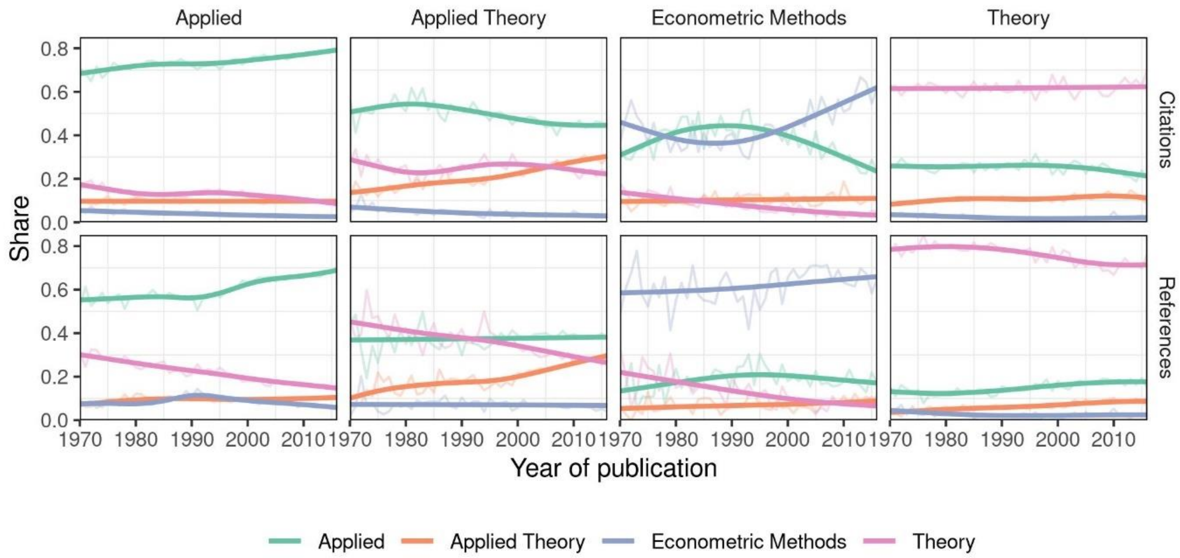
Figure 11 reveals a compelling resemblance to Figure 10, as the patterns observed in both figures align quite closely. The parallelism in the levels and trends depicted in these figures suggests a strong interaction between some fields of study other than economics and different fields of economics research.

4.3. The interplay within and between fields of economics research

In this final section, our focus shifts to examining the dynamics of citations within the field of economics. Specifically, we explore how articles within a particular field of economics research reference and are cited by articles within their own field as well as by articles from other fields. To accomplish this, we leverage the field of economics research tags assigned to each citing and referenced economics research article (see Section 3.2).

The upper panels of Figure 12 plot, for all fields of economics research, the share of citations (from economics articles) that come from a given field of economics research. For example, by 2016, nearly 80% of all citations to applied economics papers came from other applied papers. The lower panels of Figure 12 show, for all fields of economics research, the proportion of their references that cite articles of a given field of economics research. For example, by 2016, nearly 70% of all applied papers' references cited other applied papers.[35]

Figure 12. Citations and references within and between fields of economics research



Notes. Trends are smoothed by fitting a generative additive model (GAM) to the data. See Hastie, Tibshirani, and Friedman 2009). We only consider citations received within the first seven years of publication.

Figure 12 illustrates a clear tendency towards homophily in the citation behavior within different fields of economics research. This is evident in both the top and bottom panels. Homophily indicates that articles within a specific field of economics research predominantly reference and are cited by other articles within the same field. For instance, in 2016, approximately 60% of all citations received by theory articles from economics articles originated from other theory papers, and about 70% of their references to other economics articles were directed towards other theory papers. The only exception to this pattern is evident in applied theory articles, which demonstrate significant interactions with both applied and theory articles. Analyzing the four bottom panels collectively reveals that, by 2016, applied papers were consistently among the most referenced articles across all fields examined. Regardless of the field under analysis, applied papers ranked either as the most referenced or the second most

[35] The construction of this figure closely follows the methodology described for Figure 10 and Figure 11, with two key distinctions. First, it focuses on citations to and from articles within the field of economics. Second, it calculates averages across different fields of economics research instead of fields of study outside of economics. Similar to the previous figures, we only consider citations received within the first seven years since publication.

referenced articles. When examining trends over time, it becomes apparent that the inclination of articles to be cited by or to reference articles from their own field has intensified over the years, except in the case of theory. For theory articles, a relatively stable trend is observed in terms of citations, while a declining trend is observed in terms of references.

Finally, we study, for every field of economics research, which topics are cited the most by articles from different fields of economics research. For example, which are the topics present in econometric methods papers that applied papers cite most? Which are the topics present in applied papers that theory papers cite most? Etc. We begin by studying citations to applied papers. Table 9 lists, for our sample of applied papers, which are the topics whose presence correlates the most with citations coming from different fields of economics research.[36] It also presents the proportion of applied papers containing the topic, and only lists topics having a presence of at least 5% in all applied papers.

[36] We build Table 9 in a similar fashion to Table 6, but in this case, the correlations are calculated for our sample of applied papers and for citations received by every field of economics research. That is, for each topic, we calculate four correlations: one for citations coming from each field of economics research. We again consider citations received during the first seven years since publication. Only topics having a presence of at least 5% are listed.

Table 9. Topics cited the most in applied papers by articles from different fields of economics research

Citing field of economics research	Topic	Correlation with citations shares	Share of articles including the topic
Applied	effect (10.78%) yes (4.33%) control (3.06%) dummy (3%) variable (2.94%) fix (2.92%) table (2.66%) specification (2.31%)	0.206	27.72%
	sample (13.09%) estimate (9.6%) datum (4.65%) table (3.67%) report (2.41%) difference (2.1%) mean (2%) observation (1.3%)	0.126	34.47%
	age (42.35%) old (10.08%) young (8.81%) year (1.82%) effect (0.87%) datum (0.65%) decline (0.63%) rate (0.63%)	0.092	5.06%
	decline (5.67%) change (4.57%) rise (4.08%) fall (1.81%) figure (1.4%) high (1.4%) within (1.29%) percentile (1.21%)	0.083	7.14%
	treatment (21.63%) effect (3.99%) control (3.25%) treat (2.13%) difference (1.42%) experiment (1.38%) baseline (1.22%) experimental (1.05%)	0.070	6.79%
	survey (17.94%) response (8.91%) respondent (8.35%) question (8.18%) ask (4.53%) answer (3.01%) report (2.17%) wtp (1.33%)	0.066	5.27%
Applied Theory	estimate (17.06%) parameter (6.73%) estimation (3.53%) datum (2.45%) specification (2.36%) likelihood (2.09%) function (2.04%) error (1.58%)	0.129	17.81%
	price (69.94%) change (1.36%) average (0.52%) low (0.48%) high (0.38%) different (0.33%) cost (0.33%) demand (0.28%)	0.068	14.14%
	equation (54.73%) variable (8.14%) exogenous (1.92%) determine (1.38%) endogenous (1.32%) form (1.31%) equa (0.95%) substitute (0.92%)	0.065	6.21%
	cost (70.53%) low (1.34%) fix (1.33%) incur (1.03%) total (0.95%) costly (0.88%) average (0.86%) high (0.65%)	0.059	5.80%
	datum (11.65%) quarterly (7.57%) quarter (7.56%) gap (5.57%) estimate (3.95%) table (2.1%) revision (1.2%) available (1.15%)	0.043	6.14%
	time (31.07%) constant (2.28%) zero (1.37%) equal (1.14%) initial (1.03%) assumption (1.03%) section (0.96%) interval (0.9%)	0.039	5.12%
Econometric Methods	test (35.03%) power (4.77%) testing (4.35%) statistic (3.84%) alternative (2.38%) null (1.79%) sample (1.69%) regression (1.63%)	0.168	5.17%
	estimate (5.34%) estimator (3.9%) variable (3.83%) square (3.7%) estimation (3.67%) error (3.05%) least (2.72%) regression (2.21%)	0.138	7.47%
	estimate (17.06%) parameter (6.73%) estimation (3.53%) datum (2.45%) specification (2.36%) likelihood (2.09%) function (2.04%) error (1.58%)	0.119	17.81%
	equation (54.73%) variable (8.14%) exogenous (1.92%) determine (1.38%) endogenous (1.32%) form (1.31%) equa (0.95%) substitute (0.92%)	0.111	6.21%
	hypothesis (32.25%) test (11.01%) reject (8.36%) null (7.68%) datum (3.28%) significance (2.48%) alternative (1.98%) evidence (1.79%)	0.101	5.13%
	datum (11.65%) quarterly (7.57%) quarter (7.56%) gap (5.57%) estimate (3.95%) table (2.1%) revision (1.2%) available (1.15%)	0.085	6.14%
Theory	theory (13.48%) empirical (1.83%) approach (1.55%) theoretical (1.44%) view (1.21%) work (1.07%) concept (1.02%) general (1%)	0.130	7.43%
	subject (9.03%) experiment (7.56%) experimental (3.22%) table (1.79%) session (1.4%) number (1.32%) behavior (1.3%) high (1.1%)	0.094	9.34%
	cost (70.53%) low (1.34%) fix (1.33%) incur (1.03%) total (0.95%) costly (0.88%) average (0.86%) high (0.65%)	0.079	5.80%
	price (69.94%) change (1.36%) average (0.52%) low (0.48%) high (0.38%) different (0.33%) cost (0.33%) demand (0.28%)	0.048	14.14%
	market (76.39%) structure (0.97%) mar (0.64%) ket (0.52%) effect (0.37%) clearing (0.29%) kets (0.29%) different (0.28%)	0.036	6.10%
	gold (4.2%) british (3.99%) britain (3.46%) united (2.78%) london (2.48%) england (2.02%) kingdom (1.72%) great (1.25%)	0.035	5.95%

Notes. Within each citing field of economics research, topics are sorted in decreasing order according to the estimated correlations. Only topic having a presence in of at least 5% are listed. We only consider citations received within the first seven years of publication.

Table 9 shows that the topics present in applied papers that are cited the most vary greatly depending on the field of the citing articles. Applied papers citing other applied papers tend to cite articles that include topics commonly associated with causal analysis (composed of words such as *effect* and *treatment*) and data description/construction (including words such as *sample* and *survey*). Applied theory articles citing applied papers tend to cite papers including jargon commonly associated with the estimation of models' parameters (*estimate*, *parameter*, *equation*, *variable*, etc.), but they also cite a topic led by the word *price* and another led by the word *cost*. Econometric methods articles mostly cite applied papers which include topics associated with estimators (e.g., a topic led by the words *test* and *power*, and another topic led by the words *estimate* and *estimator* are listed). Theory papers tend to cite applied articles containing a topic led by the word *theory*, and, as it was the case for citations coming from applied theory papers, they also cite a topic led by the word *price* and another led by the word *cost*. Interestingly, theory papers also tend to cite applied articles containing the topic led by the word *subject* and *experiment*, suggesting that experimental economics is an interesting research topic for this field.

Table 10, Table 11, and Table 12 replicate Table 9, but focus on citations to our sample of applied theory, econometric methods, and theory articles, respectively. The consistent finding across all tables is that the topics most cited in articles from a particular field of economics research vary significantly depending on the field of the citing articles. This implies that the topics considered important or interesting within a specific field of economics research differ greatly based on the field of the citing article.

Table 10. Topics cited the most in applied theory papers by articles from different fields of economics research

Citing field of economics research	Topic	Correlation with citations shares	Share of articles including the topic
Applied	effect (10.78%) yes (4.33%) control (3.06%) dummy (3%) variable (2.94%) fix (2.92%) table (2.66%) specification (2.31%)	0.172	5.22%
	variable (9.71%) coefficient (6.31%) regression (3.75%) significant (2.25%) table (2.02%) datum (1.64%) estimate (1.15%) study (1.14%)	0.159	21.59%
	sample (13.09%) estimate (9.6%) datum (4.65%) table (3.67%) report (2.41%) difference (2.1%) mean (2%) observation (1.3%)	0.143	13.03%
	estimate (17.06%) parameter (6.73%) estimation (3.53%) datum (2.45%) specification (2.36%) likelihood (2.09%) function (2.04%) error (1.58%)	0.133	28.44%
	decline (5.67%) change (4.57%) rise (4.08%) fall (1.81%) figure (1.4%) high (1.4%) within (1.29%) percentile (1.21%)	0.077	5.08%
	industry (34.48%) manufacturing (1.9%) table (1.21%) sic (1.17%) datum (0.95%) metal (0.94%) average (0.85%) chemical (0.77%)	0.063	6.86%
Applied Theory	datum (4.18%) average (2.98%) parameter (2.86%) benchmark (2.81%) baseline (2.6%) high (2.24%) table (2.07%) figure (1.88%)	0.305	27.90%
	shock (29.43%) response (6.36%) var (1.55%) impulse (1.51%) effect (1.51%) variable (1.12%) identify (1.03%) figure (0.88%)	0.136	9.99%
	consumption (54.89%) intertemporal (3.64%) consume (3.06%) sumption (2.08%) consump (1.98%) con (1.82%) tion (0.8%) marginal (0.7%)	0.122	9.31%
	steady (18.9%) state (4.97%) transition (4.8%) economy (4.06%) initial (1.83%) dynamic (1.74%) path (1.6%) level (1.34%)	0.114	6.79%
	household (60.75%) hold (2.32%) income (2.01%) house (1.99%) survey (1.29%) total (0.8%) consumption (0.67%) include (0.52%)	0.093	6.65%
	worker (54.81%) wage (2.85%) work (1.62%) high (1.55%) job (1.11%) employ (1.01%) low (0.83%) labor (0.73%)	0.082	7.06%
Econometric Methods	estimate (17.06%) parameter (6.73%) estimation (3.53%) datum (2.45%) specification (2.36%) likelihood (2.09%) function (2.04%) error (1.58%)	0.176	28.44%
	sample (13.09%) estimate (9.6%) datum (4.65%) table (3.67%) report (2.41%) difference (2.1%) mean (2%) observation (1.3%)	0.127	13.03%
	simulation (9.04%) parameter (8.91%) simulate (4.7%) table (3.06%) mean (2.21%) draw (1.96%) method (1.82%) compute (1.76%)	0.112	8.05%
	variable (9.71%) coefficient (6.31%) regression (3.75%) significant (2.25%) table (2.02%) datum (1.64%) estimate (1.15%) study (1.14%)	0.079	21.59%
	distribution (28.53%) inequality (11.51%) mean (3.22%) gini (1.58%) measure (1.51%) equality (1.11%) distributional (0.93%) equal (0.84%)	0.072	7.37%
	equation (54.73%) variable (8.14%) exogenous (1.92%) determine (1.38%) endogenous (1.32%) form (1.31%) equa (0.95%) substitute (0.92%)	0.068	8.42%
Theory	optimal (26.27%) problem (6.15%) function (2.13%) maximize (2.12%) solution (2.08%) objective (1.4%) solve (1.37%) optimality (1.29%)	0.254	8.97%
	equilibrium (43.53%) exist (1.87%) existence (1.09%) unique (0.97%) equilib (0.94%) rium (0.87%) equi (0.82%) librium (0.68%)	0.237	8.05%
	proposition (8.83%) proof (2.72%) equilibrium (2.44%) imply (2.27%) high (2.06%) low (1.56%) decrease (1.5%) hold (1.48%)	0.190	18.76%
	steady (18.9%) state (4.97%) transition (4.8%) economy (4.06%) initial (1.83%) dynamic (1.74%) path (1.6%) level (1.34%)	0.123	6.79%
	effect (19.24%) reduce (2.83%) rise (2.79%) change (2.27%) fall (2.25%) raise (1.78%) reduction (1.52%) affect (1.43%)	0.108	24.86%
	tax (52.09%) taxis (8.95%) rate (5.2%) revenue (3.35%) taxation (2.74%) high (1.34%) base (1.33%) burden (0.95%)	0.106	7.54%

Notes. Within each citing field of economics research, topics are sorted in decreasing order according to the estimated correlations. Only topic having a presence in of at least 5% are listed. We only consider citations received within the first seven years of publication.

Table 11. Topics cited the most in econometric methods papers by articles from different fields of economics research

Citing field of economics research	Topic	Correlation with citations shares	Share of articles including the topic
Applied	sample (13.09%) estimate (9.6%) datum (4.65%) table (3.67%) report (2.41%) difference (2.1%) mean (2%) observation (1.3%)	0.197	22.31%
	estimate (17.06%) parameter (6.73%) estimation (3.53%) datum (2.45%) specification (2.36%) likelihood (2.09%) function (2.04%) error (1.58%)	0.151	30.10%
	series (16.29%) trend (13.35%) datum (5.59%) time (3.66%) estimate (2.42%) postwar (1.5%) table (1.27%) period (1.17%)	0.125	7.88%
	hypothesis (32.25%) test (11.01%) reject (8.36%) null (7.68%) datum (3.28%) significance (2.48%) alternative (1.98%) evidence (1.79%)	0.099	12.04%
	cointegrate (3.91%) phillips (3.74%) cointegration (3.68%) regression (2.29%) vector (2.01%) series (1.94%) long (1.51%) granger (1.51%)	0.085	8.22%
	price (69.94%) change (1.36%) average (0.52%) low (0.48%) high (0.38%) different (0.33%) cost (0.33%) demand (0.28%)	0.078	5.11%
Applied Theory	estimate (17.06%) parameter (6.73%) estimation (3.53%) datum (2.45%) specification (2.36%) likelihood (2.09%) function (2.04%) error (1.58%)	0.122	30.10%
	probability (44.16%) conditional (3.55%) distribution (2.08%) transition (2.06%) random (1.7%) number (1.65%) expect (1.54%) prob (1.45%)	0.084	7.74%
	price (69.94%) change (1.36%) average (0.52%) low (0.48%) high (0.38%) different (0.33%) cost (0.33%) demand (0.28%)	0.082	5.11%
	simulation (9.04%) parameter (8.91%) simulate (4.7%) table (3.06%) mean (2.21%) draw (1.96%) method (1.82%) compute (1.76%)	0.066	20.69%
	process (14.95%) conditional (5.39%) time (2.54%) series (1.73%) martingale (1.46%) mean (1.21%) journal (1.03%) stochastic (1%)	0.057	20.64%
	restriction (47.42%) restrict (12.32%) impose (8.5%) unrestricted (5.53%) deregulation (3.06%) allow (1.5%) restric (1.47%) restrictive (1.37%)	0.056	5.45%
Econometric Methods	estimator (9.44%) asymptotic (6.43%) assumption (3.12%) distribution (2.78%) sample (1.87%) asymptotically (1.81%) moment (1.34%) statistic (1.2%)	0.351	50.55%
	nonparametric (5.3%) estimation (5.15%) kernel (4.82%) function (4.3%) moment (4.06%) parametric (3.6%) semiparametric (2.81%) estimator (2.34%)	0.263	16.24%
	bound (4.34%) space (3.83%) assumption (3.28%) continuous (2.86%) compact (2.05%) convex (1.83%) function (1.82%) let (1.59%)	0.169	18.39%
	interval (6.72%) confidence (4.68%) bootstrap (3.23%) inference (1.63%) method (1.43%) sample (1.11%) estimate (1.08%) distribution (1.08%)	0.166	27.76%
	theorem (9.85%) proof (5.57%) let (5.29%) lemma (4.77%) assumption (2.6%) define (2.33%) exist (2.22%) imply (2.13%)	0.165	43.53%
	density (31.8%) distribution (13.58%) likelihood (6.38%) maximum (3.8%) censor (3.07%) observation (2.22%) normal (1.59%) function (1.56%)	0.145	12.42%
Theory	solution (9.48%) problem (5.12%) function (2.83%) convex (2.12%) concave (1.96%) point (1.82%) property (1.74%) unique (1.71%)	0.310	16.96%
	optimal (26.27%) problem (6.15%) function (2.13%) maximize (2.12%) solution (2.08%) objective (1.4%) solve (1.37%) optimality (1.29%)	0.251	6.78%
	price (69.94%) change (1.36%) average (0.52%) low (0.48%) high (0.38%) different (0.33%) cost (0.33%) demand (0.28%)	0.112	5.11%
	bound (4.34%) space (3.83%) assumption (3.28%) continuous (2.86%) compact (2.05%) convex (1.83%) function (1.82%) let (1.59%)	0.111	18.39%
	condition (48.16%) sufficient (4.68%) necessary (3.95%) must (2.15%) satisfy (1.91%) initial (1.61%) positive (1.21%) zero (1%)	0.111	10.08%
	function (45.52%) constant (2.39%) form (1.75%) respect (1.25%) derivative (1.21%) homogeneous (1.15%) derive (1.06%) partial (1.05%)	0.109	20.07%

Notes. Within each citing field of economics research, topics are sorted in decreasing order according to the estimated correlations. Only topic having a presence in of at least 5% are listed. We only consider citations received within the first seven years of publication.

Table 12. Topics cited the most in theory papers by articles from different fields of economics research

Citing field of economics research	Topic	Correlation with citations shares	Share of articles including the topic
Applied	income (67.75%) low (1.1%) high (0.86%) level (0.85%) come (0.84%) total (0.66%) equal (0.6%) disposable (0.53%)	0.088	5.17%
	country (38.49%) world (4.87%) develop (3.23%) international (3.04%) high (0.96%) development (0.84%) try (0.72%) coun (0.71%)	0.078	5.57%
	worker (54.81%) wage (2.85%) work (1.62%) high (1.55%) job (1.11%) employ (1.01%) low (0.83%) labor (0.73%)	0.077	5.59%
	wage (70.29%) high (1.18%) low (0.74%) hourly (0.67%) level (0.61%) worker (0.5%) determination (0.48%) labor (0.47%)	0.076	5.62%
	market (76.39%) structure (0.97%) mar (0.64%) ket (0.52%) effect (0.37%) clearing (0.29%) kets (0.29%) different (0.28%)	0.064	6.90%
	effect (19.24%) reduce (2.83%) rise (2.79%) change (2.27%) fall (2.25%) raise (1.78%) reduction (1.52%) affect (1.43%)	0.063	21.77%
Applied Theory	effect (19.24%) reduce (2.83%) rise (2.79%) change (2.27%) fall (2.25%) raise (1.78%) reduction (1.52%) affect (1.43%)	0.102	21.77%
	proposition (8.83%) proof (2.72%) equilibrium (2.44%) imply (2.27%) high (2.06%) low (1.56%) decrease (1.5%) hold (1.48%)	0.097	34.78%
	steady (18.9%) state (4.97%) transition (4.8%) economy (4.06%) initial (1.83%) dynamic (1.74%) path (1.6%) level (1.34%)	0.071	5.75%
	wage (70.29%) high (1.18%) low (0.74%) hourly (0.67%) level (0.61%) worker (0.5%) determination (0.48%) labor (0.47%)	0.063	5.62%
	policy (68.23%) maker (1.31%) implement (0.77%) affect (0.49%) economy (0.45%) different (0.4%) adopt (0.35%) outcome (0.31%)	0.057	5.50%
	worker (54.81%) wage (2.85%) work (1.62%) high (1.55%) job (1.11%) employ (1.01%) low (0.83%) labor (0.73%)	0.045	5.59%
Econometric Methods	function (45.52%) constant (2.39%) form (1.75%) respect (1.25%) derivative (1.21%) homogeneous (1.15%) derive (1.06%) partial (1.05%)	0.121	13.25%
	theory (13.48%) empirical (1.83%) approach (1.55%) theoretical (1.44%) view (1.21%) work (1.07%) concept (1.02%) general (1%)	0.056	15.64%
	bound (4.34%) space (3.83%) assumption (3.28%) continuous (2.86%) compact (2.05%) convex (1.83%) function (1.82%) let (1.59%)	0.051	14.30%
	theorem (9.85%) proof (5.57%) let (5.29%) lemma (4.77%) assumption (2.6%) define (2.33%) exist (2.22%) imply (2.13%)	0.049	33.33%
	income (67.75%) low (1.1%) high (0.86%) level (0.85%) come (0.84%) total (0.66%) equal (0.6%) disposable (0.53%)	0.047	5.17%
	risk (46.22%) aversion (8.21%) averse (2.17%) certainty (1.43%) risky (1.34%) absolute (1.17%) mean (0.99%) decrease (0.93%)	0.044	6.45%
Theory	game (17.82%) player (15.3%) payoff (5.12%) equilibrium (4.65%) play (4.58%) strategy (1.63%) outcome (1.23%) subgame (0.93%)	0.162	9.63%
	equilibrium (43.53%) exist (1.87%) existence (1.09%) unique (0.97%) equilib (0.94%) rium (0.87%) equi (0.82%) librium (0.68%)	0.159	27.05%
	agent (75.54%) example (0.68%) environment (0.44%) receive (0.39%) depend (0.38%) allow (0.38%) denote (0.35%) problem (0.33%)	0.127	7.90%
	payoff (13.71%) ante (4.17%) renegotiation (2.88%) outcome (1.72%) party (1.33%) suppose (1.07%) assumption (1.04%) incomplete (1.04%)	0.118	7.41%
	proposition (8.83%) proof (2.72%) equilibrium (2.44%) imply (2.27%) high (2.06%) low (1.56%) decrease (1.5%) hold (1.48%)	0.105	34.78%
	information (43.51%) inform (3.93%) know (2.82%) reveal (1.93%) uninformed (1.91%) observe (1.74%) informational (1.51%) informa (1.2%)	0.100	7.66%

Notes. Within each citing field of economics research, topics are sorted in decreasing order according to the estimated correlations. Only topic having a presence in of at least 5% are listed. We only consider citations received within the first seven years of publication.

5. Summary and conclusions

We built a large corpus of 24,273 economics research articles published in well-regarded general research economics journals from 1970 up to and including 2016. For each article, we identified its field of economics research (applied, applied theory, econometric methods, or theory). We associate fields of economics research with the methodological techniques and strategies used by articles to address a research topic or question. For each article in our sample, we collected detailed data on their citations and references. After collecting this data, we enriched it using state-of-the-art machine learning and natural language processing techniques. Having built these data, we proceeded to study trends in the publication and citation of economics research articles, focusing on documenting specialization within fields of economics research. To our knowledge, this article represents the first comprehensive attempt to document specialization within economics to such a significant extent. In doing so, we propose several innovations that extend beyond previous studies. For example, this article represents the first comprehensive attempt to combine content analysis with citation analysis to study a corpus of economics research articles.

When examining differences across fields of economics research, our findings reveal significant disparities in the content of articles, as well as in their citations and references. Regarding specialization, our analysis indicates that certain fields demonstrate an increasing trend towards specialization, while others may show the opposite. Specifically, we observe a growing specialization trend in the fields of theory and econometric methods. In contrast, applied appears to be moving in the opposite direction. Applied theory stands on the middle ground, and no clear conclusion regarding specialization in this field can be drawn. Below we expand on these results.

Both theory and econometric methods have exhibited a narrowing focus on specific research topics since the 1990s, suggesting a tendency towards specialization. The topics experiencing the most pronounced rise since the mid-1990s in theory articles are closely tied to formal mathematical proofs and game theory studies. Similarly, the topics that have shown the strongest growth in econometric methods research articles are related to computational statistics, the asymptotic properties of estimators, and estimators bounds. For the case of theory articles, a negligible rise in the share of extramural citations is observed (mainly due to citations coming from business and computer science articles), However, the share of citations coming from other fields of economics research has decreased. These two trends suggest that theory articles are becoming more specialized. For the case of econometric methods articles, both the share of extramural citations received (mainly from mathematics) and the share of citations coming from other fields of economics research have decreased. This suggests specialization as well. These patterns are accompanied by the fact that both theory and econometric methods articles are being published less frequently in general research economics journals and are receiving fewer citations compared to applied and applied theory articles.

Contrary to the previous findings, applied articles have exhibited an expansion in the diversity of topics covered since the 1990s. The topics experiencing the highest upward trends in this field are closely linked to impact analysis, causal analysis, and experimental economics. Additionally, there has been an increase in the proportion of extramural citations received by applied articles (there has been a significant rise in citations originating from disciplines such as medicine, psychology, law, and to a lesser extent, education). The topics most strongly correlated with extramural citations are predominantly present in applied papers. In addition, citations coming from other fields of economics research have increased during the period analyzed. By 2016, applied ranked among the most or second most cited field by any other field of economics research. These trends align with the increased publication frequency of applied articles in general research economics journals and their higher citation rates compared to theory articles. Overall, applied economics research seems to have become more multidisciplinary, covering a wider breadth of research topics and gaining more citations from both other fields of study and other fields of economics research.

The case of applied theory articles is somewhat in the middle. While this field has shown an expansion in the diversity of topics covered since the 1990s, this has not resulted in an increase in the proportion of extramural citations received by applied theory articles, or in the share of citations originating from other fields of economics research. In fact, applied theory articles receive the least citations from fields of study other than economics. Although citation patterns indicate a potential specialization in this field, the upward trends in the breadth of topics covered suggest the opposite, making it difficult to make a definitive statement regarding specialization in this particular field of economics research.

Scholarly articles have been shown to influence researchers' career paths, salaries, and reputations (Cole and Cole, 1967; Krampen et al., 2007; Ellison, 2013; Gibson, Anderson, and Tressler, 2014). Furthermore, they play a crucial role in shaping the rankings of departments and universities (Aguillo et al., 2010; Hazelkorn, 2015). Citation counts have now become the widely accepted standard for gauging the impact of scholarly articles, primarily due to the appeal of utilizing "*unobtrusive measures that do not require the cooperation of a respondent and do not themselves contaminate the response (i.e., they are non-reactive)*" (Smith, 1981). However, the practice of solely relying on citation analysis for comparing the impact of different scholarly articles has faced substantial criticism for its failure to consider other influential factors affecting citation patterns (see Bornmann and Daniel, 2008). One recurring criticism centers around "field-dependent factors," highlighting the variation in citation practices across different fields of study. In this study, we expand on previous literature by documenting that citation patterns vary greatly in accordance to the research strategy followed by an article, even within a field of study. Furthermore, we show that the research topics tackled by theoretical and applied economics research articles have varied greatly in the last few decades. This suggests that what constitutes applied or theoretical research in economics has evolved heavily over time, likely having an impact on citation patterns. These findings should serve as caveats when using citations as a "one-size-fits-all" yardstick to measure research outcomes, even within a field of study.

References

Aguillo, Isidro F, Judit Bar-Ilan, Mark Levene, and José Luis Ortega. 2010. "Comparing University Rankings." *Scientometrics* 85 (1): 243–56. https://doi.org/10.1007/s11192-010-0190-z.

Amami, Maha, Gabriella Pasi, Fabio Stella, and Rim Faiz. 2016. "An LDA-Based Approach to Scientific Paper Recommendation." In *Natural Language Processing and Information Systems*, edited by Elisabeth Métais, Farid Meziane, Mohamad Saraee, Vijayan Sugumaran, and Sunil Vadera, 200–210. Cham: Springer International Publishing.

Anauati, Victoria, Sebastian Galiani, and Ramiro H. Gálvez. 2016. "Quantifying the Life Cycle of Scholarly Articles across Fields of Economic Research." *Economic Inquiry* 54 (2): 1339–55. https://doi.org/10.1111/ecin.12292.

———. 2020. "Differences in Citation Patterns across Journal Tiers: The Case of Economics." *Economic Inquiry* 58 (3): 1217–32. https://doi.org/10.1111/ecin.12867.

Anderson, Katharine A, and Seth Richards-Shubik. 2022. "Collaborative Production in Science: An Empirical Analysis of Coauthorships in Economics." *The Review of Economics and Statistics* 104 (6): 1241–55. https://doi.org/10.1162/rest_a_01025.

Angrist, Josh, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. 2020. "Inside Job or Deep Impact? Extramural Citations and the Influence of Economic Scholarship." *Journal of Economic Literature* 58 (1): 3–52. https://doi.org/10.1257/jel.20181508.

Angrist, Joshua, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu. 2017. "Economic Research Evolves: Fields and Styles." *American Economic Review* 107 (5): 293–97. https://doi.org/10.1257/aer.p20171117.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30. https://doi.org/10.1257/jep.24.2.3.

Backhouse, Roger, and Beatrice Cherrier. 2014. "Becoming Applied: The Transformation of Economics after 1970." The Center for the History of Political Economy Working Paper Series No. 2014-15. http://dx.doi.org/10.2139/ssrn.2526274.

Becker, Gary S, and Kevin M Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." *The Quarterly Journal of Economics* 107 (4): 1137–60. https://doi.org/10.2307/2118383.

Beltagy, Iz, Kyle Lo, and Arman Cohan. 2019. “SciBERT: A Pretrained Language Model for Scientific Text.” In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3615–20. Hong Kong, China: Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1371.

Biddle, Jeff E, and Daniel S Hamermesh. 2017. “Theory and Measurement: Emergence, Consolidation, and Erosion of a Consensus.” *History of Political Economy* 49 (Supplement): 34–57. https://doi.org/10.1215/00182702-4166251.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. “Latent Dirichlet Allocation.” *The Journal of Machine Learning Research* 3: 993–1022.

Bornmann, Lutz, and Hans-Dieter Daniel. 2008. “What Do Citation Counts Measure? A Review of Studies on Citing Behavior.” *Journal of Documentation* 64 (1): 45–80.

Card, David, and Stefano DellaVigna. 2013. “Nine Facts about Top Journals in Economics.” *Journal of Economic Literature* 51 (1): 144–61. https://doi.org/10.1257/jel.51.1.144.

Chiappori, Pierre-André, and Steven D Levitt. 2003. “An Examination of the Influence of Theory and Individual Theorists on Empirical Research in Microeconomics.” *American Economic Review* 93 (2): 151–55. https://doi.org/10.1257/000282803321946967.

Cole, Stephen, and Jonathan R Cole. 1967. “Scientific Output and Recognition: A Study in the Operation of the Reward System in Science.” *American Sociological Review*, 377–90.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–86. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.

Dirac, Paul Adrien Maurice. 1940. “The Relation between Mathematics and Physics.” *Proceedings of the Royal Society of Edinburgh* 59: 122–129. https://doi.org/10.1017/S0370164600012207.

Duhem, Pierre. 1976. “Physical Theory and Experiment.” In , edited by Sandra G Harding, 1–40. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-1863-0_1.

Dunham, William. 1994. *The Mathematical Universe: An Alphabetical Journey Through the Great Proofs, Problems, and Personalities*. New York: Wiley.

Ellison, Glenn. 2013. “How Does the Market Use Citation Data? The Hirsch Index in Economics.” *American Economic Journal: Applied Economics* 5 (3): 63–90. https://doi.org/10.1257/app.5.3.63.

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. “The Superiority of Economists.” *Journal of Economic Perspectives* 29 (1): 89–114. https://doi.org/10.1257/jep.29.1.89.

Galiani, Sebastian, and Ramiro H Gálvez. 2019. “An Empirical Approach Based on Quantile Regression for Estimating Citation Ageing.” *Journal of Informetrics* 13 (2): 738–50. https://doi.org/https://doi.org/10.1016/j.joi.2019.03.014.

Gálvez, R.H. 2017. “Assessing Author Self-Citation as a Mechanism of Relevant Knowledge Diffusion.” *Scientometrics* 111 (3). https://doi.org/10.1007/s11192-017-2330-1.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. “Text as Data.” *Journal of Economic Literature* 57 (3): 535–74. https://doi.org/10.1257/jel.20181020.

Gibson, John, David L Anderson, and John Tressler. 2014. “Which Journal Rankings Best Explain

Academic Salaries? Evidence from the University of California." *Economic Inquiry* 52 (4): 1322–40. https://doi.org/https://doi.org/10.1111/ecin.12107.

Hall, David, Daniel Jurafsky, and Christopher D Manning. 2008. "Studying the History of Ideas Using Topic Models." In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 363–71. Honolulu, Hawaii: Association for Computational Linguistics. https://aclanthology.org/D08-1038.

Hamermesh, Daniel S. 2013. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature* 51 (1): 162–72. https://doi.org/10.1257/jel.51.1.162.

———. 2018. "Citations in Economics: Measurement, Uses, and Impacts." *Journal of Economic Literature* 56 (1): 115–56. https://doi.org/10.1257/jel.20161326.

Harrow, Keith. 1979. "Theoretical and Applied Computer Science: Antagonism or Symbiosis?" *The American Mathematical Monthly* 86 (4): 253–60. http://www.jstor.org/stable/2320741.

Hastie, T, R Tibshirani, and J H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York, NY, USA: Springer.

Hazelkorn, Ellen. 2015. *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. Springer.

Heckman, James J, and Sidharth Moktan. 2020. "Publishing and Promotion in Economics: The Tyranny of the Top Five." *Journal of Economic Literature* 58 (2): 419–70. https://doi.org/10.1257/jel.20191574.

Hu, Diane J. 2009. "Latent Dirichlet Allocation for Text, Images, and Music." *University of California, San Diego*. http://cseweb.ucsd.edu/~dhu/docs/research_exam09.pdf.

Jurafsky, Daniel, and James H Martin. 2009. *Speech and Language Processing (2nd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Koch, Sigmund. 1973. "Theory and Experiment in Psychology." *Social Research* 40 (4): 691–707. http://www.jstor.org/stable/40970161.

Krampen, Günter, Ralf Becker, Ute Wahner, and Leo Montada. 2007. "On the Validity of Citation Counting in Science Evaluation: Content Analyses of References and Citations in Psychological Publications." *Scientometrics* 71 (2): 191–202. https://doi.org/10.1007/s11192-007-1659-2.

Krishnan, Armin. 2009. "What Are Academic Disciplines? Some Observations on the Disciplinarity vs. Interdisciplinarity Debate." NCRM Working Paper Series. National Centre for Research Methods. http://eprints.ncrm.ac.uk/783/1/what_are_academic_disciplines.pdf.

Levitt, Steven D. 1998a. "Juvenile Crime and Punishment." *Journal of Political Economy* 106 (6): 1156–85. https://doi.org/10.1086/250043.

Levitt, Steven D. 1998b. "Why Do Increased Arrest Rates Appear to Reduce Crime: Deterrence, Incapacitation, or Measurement Error?" *Economic Inquiry* 36 (3): 353–72. https://doi.org/https://doi.org/10.1111/j.1465-7295.1998.tb01720.x.

Longo, Giuseppe, and Ana M Soto. 2016. "Why Do We Need Theories?" *Progress in Biophysics and Molecular Biology* 122 (1): 4–10. https://doi.org/https://doi.org/10.1016/j.pbiomolbio.2016.06.005.

Mailath, George J, Larry Samuelson, and Jeroen M Swinkels. 1993. "Extensive Form Reasoning in Normal Form Games." *Econometrica* 61 (2): 273–302. https://doi.org/10.2307/2951552.

Mustard, David B. 2003. "Reexamining Criminal Behavior: The Importance of Omitted Variable Bias." *The Review of Economics and Statistics* 85 (1): 205–11.

https://doi.org/10.1162/rest.2003.85.1.205.

Neff, Bryan D., and Julian D. Olden. 2010. “Not So Fast: Inflation in Impact Factors Contributes to Apparent Improvements in Journal Quality.” *BioScience* 60 (6): 455–59. https://doi.org/10.1525/bio.2010.60.6.9.

Panhans, Matthew T, and John D Singleton. 2017. “The Empirical Economist’s Toolkit: From Models to Methods.” *History of Political Economy* 49 (Supplement): 127–57. https://doi.org/10.1215/00182702-4166299.

Phillips, Colin, Phoebe Gaston, Nick Huang, and Hanna Muller. 2021. “Theories All the Way Down: Remarks on ‘Theoretical’ and ‘Experimental’ Linguistics.” In *The Cambridge Handbook of Experimental Syntax*, edited by GrantEditor Goodall, 587–616. Cambridge Handbooks in Language and Linguistics. Cambridge University Press. https://doi.org/10.1017/9781108569620.023.

Pieters, Rik, and Hans Baumgartner. 2002. “Who Talks to Whom? Intra- and Interdisciplinary Communication of Economics Journals.” *Journal of Economic Literature* 40 (2): 483–509. https://doi.org/10.1257/002205102320161348.

Poincaré, Henri. 1902. “Relations Between Experimental Physics and Mathematical Physics.” *The Monist* 12 (4): 516–43. http://www.jstor.org/stable/27899347.

Putnam, Hilary. 1979. “The ‘Corroboration’ of Theories.” In *Mathematics, Matter and Method: Philosophical Papers*, 2nd ed., 1:250–269. Cambridge University Press. https://doi.org/10.1017/CBO9780511625268.018.

Ramaley, Francis. 1930. “Specialization in Science.” *Science* 72 (1866): 325–26. https://doi.org/10.1126/science.72.1866.325.

Reeves, Scott, Mathieu Albert, Ayelet Kuper, and Brian David Hodges. 2008. “Why Use Theories in Qualitative Research?” *BMJ* 337. https://doi.org/10.1136/bmj.a949.

Řehůřek, Radim, and Petr Sojka. 2010. “Software Framework for Topic Modelling with Large Corpora.” In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA.

Smith, Linda C. 1981. “Citation Analysis.” *Library Trends* 30 (1): 83–106. http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ284261.

Walsh, John P, and Nancy G Maloney. 2007. “Collaboration Structure, Communication Media, and Problems in Scientific Work Teams.” *Journal of Computer-Mediated Communication* 12 (2): 712–32. https://doi.org/https://doi.org/10.1111/j.1083-6101.2007.00346.x.

Wen, Quan. 1994. “The ‘Folk Theorem’ for Repeated Games with Complete Information.” *Econometrica* 62 (4): 949–54. http://www.jstor.org/stable/2951740.

———. 2002. “A Folk Theorem for Repeated Sequential Games.” *The Review of Economic Studies* 69 (2): 493–512. https://doi.org/10.1111/1467-937X.00214.

Wray, K Brad. 2005. “Rethinking Scientific Specialization.” *Social Studies of Science* 35 (1): 151–64. http://www.jstor.org/stable/25046633.

Ziman, J M. 1987. *Knowing Everything about Nothing: Specialization and Change in Research Careers*. Cambridge University Press. https://books.google.com.ar/books?id=ZEaCTDOw-N8C.