ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Brodeur, Abel; Bruns, Stephan B.; Herwartz, Helmut; Islam, Chris-Gabriel

Working Paper

How did empirical economic research improve over the last decades: Recent trends and prospects

I4R Discussion Paper Series, No. 315

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Brodeur, Abel; Bruns, Stephan B.; Herwartz, Helmut; Islam, Chris-Gabriel (2026) : How did empirical economic research improve over the last decades: Recent trends and prospects, I4R Discussion Paper Series, No. 315, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/342985

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

INSTITUTE for REPLICATION

No. 315
I4R DISCUSSION PAPER SERIES

How did empirical economic research improve over the last decades: Recent trends and prospects

Abel Brodeur

Stephan B. Bruns

Helmut Herwartz

Chris-Gabriel Islam

August 2026

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 315

How did empirical economic research improve over the last decades: Recent trends and prospects

Abel Brodeur[1,2,3], Stephan B. Bruns[4,5,6], Helmut Herwartz[7], Chris-Gabriel Islam[4,7]

[1]University of Ottawa/Canada
[2]Institute for Replication
[3]IZA – Institute for Labor Economics, Bonn/Germany
[4]Hasselt University, Hasselt/Belgium
[5]Meta-Research Innovation Center at Stanford (METRICS), Stanford/USA
[6]International Center for Higher Education Research (INCHER), Kassel/Germany
[7]Georg August University of Göttingen/Germany

AUGUST 2026

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the ZBW – Leibniz Information Centre for Economics, and RWI – Leibniz Institute for Economic Research, I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur — ***University of Ottawa***
Jörg Ankel-Peters — ***RWI – Leibniz Institute for Economic Research***

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

How did empirical economic research improve over the last decades: Recent trends and prospects*

Abel Brodeur[1,2,3], Stephan B. Bruns[4,5,6,†], Helmut Herwartz[7] and Chris-Gabriel Islam[4,7,8]

[1]University of Ottawa, Canada
[2]Institute for Replication
[3]IZA – Institute of Labor Economics, Germany
[4]Hasselt University, Belgium
[5]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, USA
[6]International Center for Higher Education Research (INCHER) Kassel, Germany
[7]Georg August University of Göttingen, Germany
[†]Corresponding author: stephan.bruns@uhasselt.be

August 17, 2026

Abstract

We map the evolution of economics research from 1998 to 2019 to assess progress and identify remaining challenges. Using text mining of 30,675 full-text articles and an in-depth analysis of 578,132 statistical tests from 3,746 articles across a broad range of journals, we document changes in research practices over time. We highlight the influence of the credibility revolution and the open science movement. Awareness of reproducibility has increased modestly, largely driven by journal adoption of data and code sharing policies. At the same time, reliance on non-public data has grown, especially among authors at top-five universities, posing new challenges for transparency. Sample sizes have increased substantially, while reported effect sizes have declined, consistent with reduced exaggeration and potentially more policy-relevant estimates. Econometric methods and causal inference practices have become more sophisticated, and articles now report more tables and statistical tests. However, pre-analysis plans, power analyses accompanying null results, and corrections for multiple hypothesis testing remain uncommon. Article length has remained stable or declined slightly, and authors from top-five universities continue to represent a

*Acknowledgements: We thank two anonymous reviewers, the editor and participants at the MAER-Net conference 2024 in Augsburg for their helpful comments. Stephan Bruns, Helmut Herwartz and Chris-Gabriel Islam are grateful for funding from the German Research Foundation (DFG) under the project "Replications in Empirical Economics: Necessity, Incentives and Impact" with follow up project "Selective reporting and the evolving research landscape in economics" (project number: 405039391). Stephan Bruns is also grateful for funding from Research Foundation Flanders (FWO) and Fund for Scientific Research (FNRS) under the project "Evidence-based finance: Empirical analysis of p-hacking and advancement of methods to promote improved decision making" (project numbers: G0A4G24N and TW01523F) and Chris-Gabriel Islam is grateful for funding from Hasselt University within the framework of BOF-BILA (BOF21BL08). The replication package with data and code is available at: https://github.com/Irazall/replication-package-landscape/tree/main

large and stable share of publications in top journals. Our findings provide a baseline for tracking future developments and for guiding efforts to further strengthen the credibility and transparency of economics research.

Keywords: Replication crisis; Credibility revolution; Open science; Trends; Empirical research
JEL codes: A19, C18, C40, C80

1 Introduction

Economics research has become increasingly empirical over the last decades (Hamermesh 2025; Mueller-Langer et al. 2019; Hamermesh 2013). The first empirical wave began in the late 1970s, driven by the availability of advanced computing power. This advancement simplified the estimation of econometric models based on observational data. Estimating more econometric models in less time also facilitated the usage of questionable research practices, such as searching for statistically significant specifications and then selectively reporting only these specifications, which is also known as p-hacking (Lybbert and Buccola 2021). The misuse of statistical significance and the lack of proper identification of causal effects drew the attention of many critics (e.g., Kennedy 2002; McCloskey and Ziliak 1996; Leamer 1983). In the 2000s, economics responded with the second empirical wave, also known as the credibility revolution (Lybbert and Buccola 2021; Angrist and Pischke 2010). This movement aims to improve econometric methods and focuses more on experimental and quasi-experimental studies to identify causal effects. In the 2010s, the replication crisis gained momentum in economics and calls for open science became louder, trying to make research more transparent and aiming to improve general research practices (Brodeur et al. 2016; Bruns and Ioannidis 2016; Camerer et al. 2016; Collaboration 2015).

We map the development of economics research between 1998 and 2019 to take stock and identify areas for further improvement. We structure our analysis by shedding light on two significant trends: the credibility revolution and the open science movement caused by the replication crisis. Additionally, we provide insights on general trends in the field of economics.

Our analysis is descriptive and contributes to the literature on trends in economics research in three ways. First, we use text mining for the full text of more than 30,000 published articles from a broad range of economics journals, providing a comprehensive and inclusive overview on trends in economics research. Previous research on trends usually focuses on a few top journals or prestigious working paper series (e.g., Hamermesh 2025; Garg and Fetzer 2025; Goldsmith-Pinkham 2024; Currie et al. 2020; Card and DellaVigna 2013; Hamermesh 2013) and the insights from this research might not generalize to other outlets. Our analysis, therefore, allows us to take stock of economics research more broadly. Second, we also add to the literature by not only analyzing text but by providing an in-depth analysis of 578,132 statistical tests from 3,746 articles. This allows us to shed light on how reporting of statistical findings changes over time, including trends in sample sizes, statistical significance, and (exaggerated) effect sizes. Third, we provide a comprehensive assessment of the open science movement and replication crisis. Such an assessment is available in psychology (Hardwicke et al. 2022) and the social sciences (Hardwicke et al. 2020), but is lacking for economics. We empirically add to the extensive literature review by Christensen and Miguel (2018). Given the high policy relevance of economics research, taking stock of its reproducibility and transparency practices appears to be of particular importance and our analysis can be used as a baseline to track future developments.

General trends

First, we analyze general trends in economics, such as the size and gender composition of author teams. Earlier studies indicate a rise in the number of authors in the top-five journals (Card and DellaVigna 2013; Hamermesh 2013). Evidence for gender bias in economics is mixed, with some suggesting that female authors might be held to higher

standards (Bruns et al. 2025; Card et al. 2020). Previous research also analyzed the balance between empirical and theoretical articles, as well as the range of research topics (Mueller-Langer et al. 2019; Card and DellaVigna 2013; Hamermesh 2013).

Credibility revolution

The second trend that we analyze refers to the credibility revolution. Following up on the critique by Leamer (1983) about the missing experimental character of economics in the 1980s, Angrist and Pischke (2010) come to the conclusion that most research areas in economics experienced a credibility revolution. Due to more and better data and better research designs, but also due to less articles with fundamental discussions about methods, economics became more empirical and more focused on causality. Experimental and quasi-experimental research designs and methods for identification are steadily being added to the economics toolkit and the number of causal claims is rising (Garg and Fetzer 2025). Instead of replacing more traditional methods, quasi-experimental methods are used as additional empirical evidence. One aspect of the credibility revolution is that authors no longer rely on a single method but rather use a variety of methods, which is coined as the collage approach (Currie et al. 2020).[1] In addition to the focus on causality, another trend within the credibility revolution in economics is the improvement of econometric methods, e.g., the widespread usage of robust and clustered inference or non-parametric and semi-parametric regression (Stock 2010).

Open science movement

The third trend that we analyze refers to the open science movement. The initial spark for this movement was caused by the replication crisis which has attracted attention for almost two decades and is prevalent in most disciplines, from natural sciences to social sciences (Sayre and Riegelman 2018). In economics, about 85 % of the published claims from 101 papers in leading journals were computationally reproducible, that is, the published claims can be reproduced with available code and data (Brodeur et al. 2026). Computational reproducibility is higher in economics and political science compared to other social sciences, mostly because it is more common to publish data and code (Miske et al. 2026). When published claims are exposed to robustness analyses, on average 72 % of the claims remained robust, but with variation ranging from 45 % to 87 %, depending on the type of robustness check (Brodeur et al. 2026). When new independent analyses were conducted for the same research questions, potentially involving new data, only about 50 % of published claims are robust in the social and behavioral sciences (Tyner et al. 2026). Over the last decades, the meta-analysis literature persistently documented the presence of selective reporting in economics, such as p-hacking and publication bias (e.g., Stanley 2001; Doucouliagos and Paldam 2008; Bruns et al. 2014; Imai et al. 2021; Bartoš et al. 2024; Cala et al. 2026). These findings are corroborated by the larger meta-research literature that diagnoses the presence of selective reporting by analyzing discontinuities in the distribution of published z- and p-values (e.g., Brodeur et al. 2016; Bruns et al. 2019; Vivalt 2019; Elliott et al. 2022; Bruns et al. 2024a). In addition to selectively reporting results, a large-scale study revealed that around 15 % of the articles published in 30 top economics journals contain at least one strong statistical reporting error in which an eye-

[1] This term is related to a similar term coined by Munafò and Smith (2018): triangulation.

catcher labels a finding as statistically significant, but the reported statistical values do not imply statistical significance or *vice versa* (Bruns et al. 2023).[2]

One crucial reason for the surge of this crisis could be that leading journals prioritize novelty over the robustness of findings (Moody et al. 2022), resulting in incentives for questionable research practices and fraud. Most disciplines responded to the replication crisis by promoting open science. Open science seeks to ensure that scientific knowledge is freely available, easily accessible, and reusable by all, thereby fostering collaborations and information sharing to advance science and society (UNESCO 2021). The most important characteristics of open science at the article level are open data, open code, and pre-registration. Since 2005, the number of economics journals mandating the availability of data and code has been constantly increasing (e.g., Bruns et al. 2023; Askarov et al. 2023; Brodeur et al. 2024a; Vlaeminck 2021; Mueller-Langer et al. 2019).[3] Compliance with this type of policy has been frequently questioned in the literature (e.g., Brodeur et al. 2024a; Askarov et al. 2023; Christensen et al. 2019). While an empirical assessment of transparency and reproducibility practices in psychology and the social sciences revealed that sharing raw data, sharing analysis scripts, pre-registering studies, and conducting replications are still rare to non existent (Hardwicke et al. 2022; Hardwicke et al. 2020), a survey among authors publishing in top economics journals as well as PhD students from the top 20 North American Economics departments suggests that transparency and reproducibility practices may be on the rise in economics, at least at the top (Swanson et al. 2020; Miguel 2021).

Our general findings suggest that between 1998 and 2019, economics research has become more empirical, more female-authored, and less single-authored. Male researchers have remained more common than female researchers in economics, and authors from top-five universities have represented a substantial fraction of top publications. While article length in top-five journals is increasing, article length in non top-five journals has remained stable or even decreased. Identifying causal effects is key in economics, with difference-in-differences and instrumental variables being widely used approaches, while RCTs have become more important. Econometric methods have improved steadily, e.g., by using clustered inference, while more attention to the correction for multiple hypothesis testing is needed. We also document more tables and tests being presented in each article, potentially suggesting that economists perform and include more robustness checks in their articles. The awareness of reproducibility and open science has increased slightly, but the main agents of the open science movement seem to be journals implementing data and code availability policies. However, the use of non-public data has increased, especially among authors from top-five universities, posing a challenge for open science. Economics research does not distinguish between exploratory and hypothesis-testing research, with little emphasis on pre-analysis plans, and it is still uncommon — though increasing — to publish null effects. Sample sizes have increased, but instead of decreasing p-values, we observe that reported effect sizes have decreased.

In the next section, we will provide an overview of our data and our methods. Sections 3 to 5 provide results of the analysis on general trends, the credibility revolution,

[2]In addition, a business paper from 1981 addresses fundamental problems in replicating research results (Reid et al. 1981), indicating that concerns about this issue existed in the field long before the present crisis.

[3]In 2004, the *American Economic Review* and *Econometrica* announced a data-and-code-sharing policy. The first economics journal to implement such a policy was the *Journal of Money, Credit and Banking* in 1982.

and the open science movement, respectively. The last section discusses the results and concludes. For a more detailed description of our data and methods, we refer to the Online Appendix in the next sections.

2 Data and methods

2.1 Data

In order to map recent trends in economics, we focus primarily on published articles. An article consists of two main elements: (1) the full text, including title, abstract, and section headlines, and (2) its statistical tests.

Our first data source is compiled by DORIS, which is a software developed by Bruns et al. (2023). The data include the full text of 30,675 articles and detailed information on 578,132 statistical tests (from 3,746 articles) from more than 30 top economic journals that provide their articles as HTML documents.

Using data provided by DORIS, we build three samples (cf. Table 1). Our first sample (*all articles: 31k*) comprises all articles scraped by DORIS published between 1998 and 2019.[4] This results in the availability of full text for 30,675 articles from 33 journals. These journals resemble the top 50 economic journals based on IDEAS/RePEc Ranking Simple Impact Factors (Last 10 Years) as of December 2018 that offer their articles as HTML documents. The sample also contains articles from *Research Policy*.[5] As journals vary in the start of publishing their articles in HTML, some journals in our sample start later than 1998. For text mining and analyzing general trends, we mainly use the sample of all articles.

Our second sample (*empirical articles: 22k*) is a subsample of all articles. It contains articles that we identify as empirical papers, that is, articles must contain at least one table and mention either the word "data" or the word "regression" at least once (cf. Section 2.2 for information on the retrieval of the relevant information). Our definition of empirical articles is almost congruent with previous literature (Mueller-Langer et al. 2019).

Our third sample (*empirical articles with statistical information: 4k*) is a subsample of the empirical articles and contains all articles for which DORIS is able to interpret statistical tests. DORIS collects statistical tests that are reported using regression coefficients along with standard errors or t-, z- or p-values. Additionally, DORIS collects eye-catchers and the maximum sample size per table. DORIS cannot extract statistical tests that report coefficients together with confidence intervals or statistical tests from journals that prohibit eye-catchers (e.g.,*Journal of Business & Economic Statistics*).[6] Statistical tests are also not extracted if the full text mentions one-sided tests. The performance of DORIS in correctly identifying statistical tests was evaluated on around 3,200 statistical tests from 32 journals. The false discovery rate is low with 1.2 % (Bruns et al. 2023). This results in

[4]In 2016, the number of accessible journals drops substantially to 19 journals due to the dispute between the publisher *Elsevier* and the German universities. For more information, consult, e.g., https://www.science.org/news/2016/12/thousands-german-researchers-set-lose-access-elsevier-journals, last retrieved on 31.08.2024.

[5]The sample does not include journals from *Springer* as a publisher (*Journal of Economic Growth*, *IMF Economic Review* and *Experimental Economics*) even though these articles are also available in HTML. Bruns et al. (2023) argue that *Springer*'s conditions to scrape these journals were not acceptable for a scientific project.

[6]Journals of the *American Economic Association* prohibit eye-catchers as well, but are not part of our sample as they are not available in HTML.

Table 1: Overview about the samples

Sample	Number of articles	Number of journals
All articles: 31k	30,675	33
Empirical articles: 22k	21,775	32
Empirical articles with statistical information: 4k	3,746	31
Articles with data and/or code: 11k	11,140	17
Matched articles with data and/or code: 1.5k	1,496	5

3,746 articles from 31 journals with a total of 172,040 main tests and 406,092 non-main tests.[7] Unlike the samples *all articles* and *empirical articles*, this sample ends in 2016. The sample *empirical articles with statistical information* is used mainly to determine trends with respect to statistical significance, the number of observations, or effect sizes.

In addition, we use data that stem from the software *Find Economic Articles with Data*[8] for the fourth and fifth samples. The fourth sample (*articles with data and/or code: 11k*) contains 11,140 articles with published data and/or code from 17 journals. Data availability and the file size of the replication package are also reported for each article. We further matched the journals and years of the fourth data set with the data set *empirical articles* resulting in five journals starting in 2011. This fifth data set (*matched articles with data and/or code: 1.5k*) is reduced to five journals as the overlap between the 32 journals of *empirical articles* and the 17 journals of *articles with data and/or code* is limited, as the latter contains seven journals from the *American Economic Association* that are not available in HTML. In the five overlapping journals, we identified 1,496 empirical articles of which 829 articles provide data and/or code, as identified by the software *Find Economic Articles with Data*.[9] In this fifth data set, all articles were published under an active data and/or code availability policy.

A detailed overview of the number of articles per journal per year per sample is given in Tables B.1 to B.5 in the Online Appendix.

2.2 Methods

This study uses text mining techniques for the full text of 30,675 articles provided by DORIS. In order to flexibly search for keywords, we use regular expressions. The particular regular expressions used are listed in Table B.6 in the Online Appendix. Before applying the regular expressions, we preprocess the text of the HTML documents. To extract title, abstract, headlines, and full text, we use the Python package *beautifulsoup4*, which allows us to retrieve information based on the HTML tag. By doing so, we are able to extract the information from the full text, including headlines, titles, and abstract, but exclude tables and figures, as well as references.

[7] DORIS categorizes a test as a main test if it appears in the first three rows of a table that is neither a robustness check nor part of the appendix. This classification is based on discontinuities in the distribution of z-values around significance thresholds (Bruns et al. 2023).

[8] Cf. https://ejd.econ.mathematik.uni-ulm.de/ provided by Ulm University (data retrieved on 29.09.2023).

[9] We cannot rule out that some articles that provide data and/or code may have been missed by the software.

Metadata on the articles, e.g., JEL classes and authors, are collected by DORIS during the scraping process. In case of missing affiliations, we coded the affiliations manually. Using regular expressions, we detect whether the respective affiliations contain one of the top 25 universities of the respective publishing year. The top universities for each year are gathered from the Economic Institution Rankings at IDEAS.[10] The gender of the authors is detected via the dictionary-based Python package *gender-guesser*. We set the detector to have a higher specificity in the direction of detecting female authors, which implies that we might underestimate the share of female authors. For the extraction of the software used in an article, we use the module *software-mentions* from the software *GROBID* (v0.7.3, cf. GROBID 2023) which is based on machine learning. As this module cannot read HTML documents, we use the respective PDFs as a data source. When analyzing derived quantities, e.g., Partial Correlation Coefficient (PCC), the explanation and formula are given directly in the respective section.

The matching between data from DORIS and data from Ulm University is performed using fuzzy string matching and the Levenshtein distance. In order to guarantee a high quality of the matching, one author manually controlled the matching and deleted false matchings.

3 General trends

3.1 Mostly stable research topics

Research topics in economics research are commonly categorized using JEL codes. The last revision was initiated in 1988 by John Pencavel, then editor of the Journal of Economic Literature (Cherrier 2017). The smallest categories in our sample of all articles are A (General Economics and Teaching: 0.12 %), B (History of Economic Thought: 0.11 %) and Y (Miscellaneous Categories: 0.01 %) whereas C (Mathematical and Quantitative Methods: 15.50 %) and G (Financial Economics: 16.56 %) resemble the largest categories. We find that the share of most primary JEL categories remains stable, with some trends possibly influenced by recent events (cf. Figure 1). The highest average increase per year is attributed to category G, financial economics, with 0.56 %. A two-sample t-test with $p = 0.002$ suggests that the breakpoint is possibly in 2008, having a mean share per year of 13.74 % until 2007 and a mean share per year of 19.51 % from 2008. Hence, the financial crisis of 2008 could have promoted the increase in financial studies. The highest average decrease per year is attributed to category F (International Studies) with 0.47 % per year. Additionally, category C (Mathematics and Quantitative Methods) shows a high average decrease of 0.32 % per year. This reflects the trend toward less fundamental theory and more empirical and experimental studies (Hamermesh 2013; Angrist and Pischke 2010). Note that trend changes originating after 2016 need to be interpreted with caution as Elsevier journals drop out of the sample. Overall, the results indicate that most topics in our sample, i.e., JEL codes, remain rather stable. However, using a different classification that is based not only on JEL codes but also on the titles, keywords, the publishing journal, and the fields of the articles cited by the article, it has been shown that microeconomic topics increased (Angrist et al. 2017).

[10] Cf. https://ideas.repec.org/top/top.inst.all.html, last retrieved on 27.12.2024.

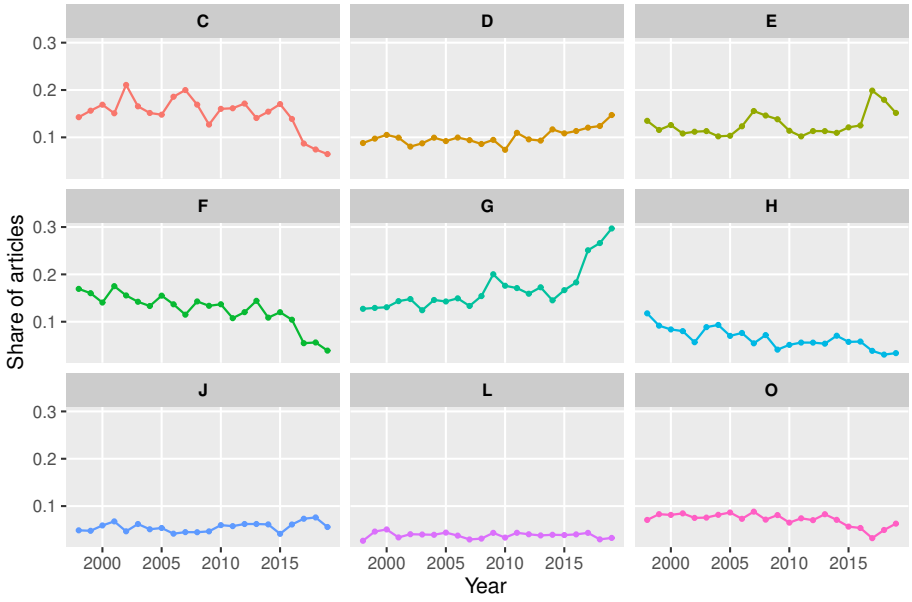


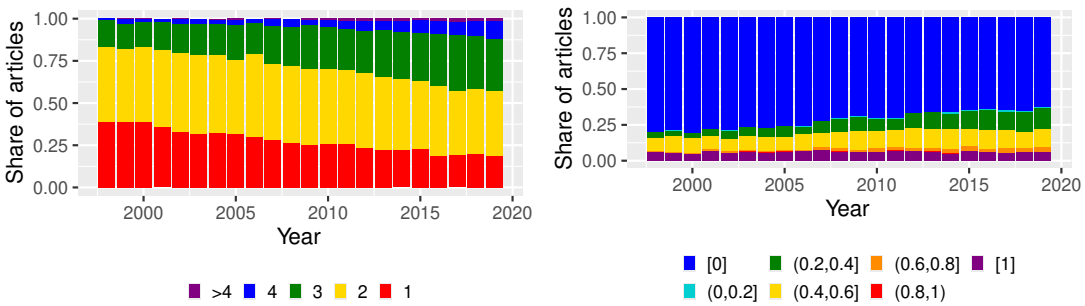
Figure 1: JEL codes per year (all articles: 31k)

Notes: We focus on JEL codes with at least a share of 5 % in one of the years for the sake of illustration. C - Mathematical and Quantitative Methods, D - Microeconomics, E - Macroeconomics and Monetary Economics, F - International Economics, G - Financial Economics, H - Public Economics, J - Labor and Demographic Economics, L - Industrial Organization, O - Economic Development.

3.2 Larger and more female author teams

Compared to other disciplines, economics is a discipline with fewer authors per article (Islam 2024). However, there have been notable trends regarding the number of authors in the last decades (cf. Figure 2a). The share of single-authored papers decreased from around 40 % in the late 1990s to around 20 % in the late 2010s, leaving single-authored articles still an important type of article in economics. Note that this persistent share of single-authored articles might be due to the tradition of job market papers in economics. At the same time, the share of articles with four or more authors increase to around 10 %, underlining the importance of collaboration between researchers while facing an increasing degree of complexity in economics research (Hamermesh 2013). Our findings are consistent with previous literature (Hamermesh 2025; Fire and Guestrin 2019; Card and DellaVigna 2013; Hamermesh 2013).

In addition, we analyze the share of female authors per year (cf. Figure 2b). We see that all-female author teams remain stable over time with around 6 % while all-male author teams decrease from around 80 % in 1998 to around 60 % in 2019. These numbers align with previous findings in the literature for the *American Economic Review*, the *Journal of Political Economy*, and the *Quarterly Journal of Economics* (Hamermesh 2013). Our findings suggest that economics research has become more feminine in recent years, but male researchers remain more common than female researchers.



(a) Number of authors per year (b) Share of female authors per year

Figure 2: Trends related to author teams (all articles: 31k)

Notes: Figures show (a) number of authors per year and (b) share of female authors per year.

3.3 Increasing share of empirical articles

It is well known that the number of published articles and journals has increased tremendously in the last few decades in all disciplines (Fire and Guestrin 2019). The number of articles in our *all articles* sample increases from 626 in 1998 to 2,090 in 2016. First, more and more journals started to publish their articles in HTML. Second, journals such as *Research Policy* or *The Review of Financial Studies* drastically increased the number of published articles per year (cf. Table B.1 in the Online Appendix for details). From 1998 to 2019, the share of empirical articles rises from around 57 % to 80 % (Figure 3). The rise can be explained by the increased availability of data, often in real-time (Einav and Levin 2014), computational advances, and the focus on identifying causality (Lybbert and Buccola 2021). Our findings are in the same ballpark as suggested in previous literature that used smaller or different samples (Mueller-Langer et al. 2019; Angrist et al. 2017; Hamermesh 2013).

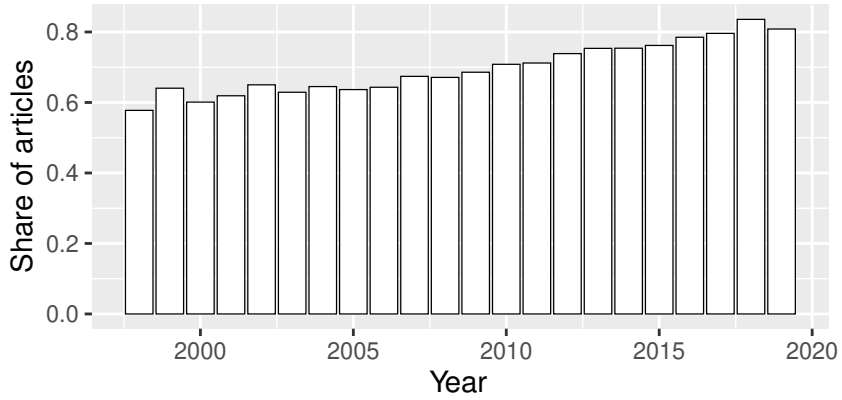


Figure 3: Share of empirical articles per year (all articles: 31k)

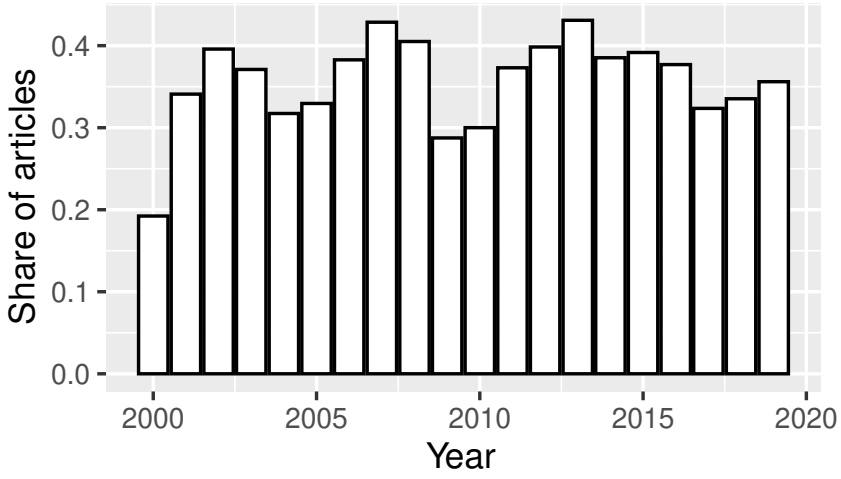


Figure 4: Share of articles in top journals per year with at least one affiliation to a top-five university (empirical articles: 22k)

Notes: Top journals in this figure are defined as *Quarterly Journal of Economics, Journal of Political Economy* and *Review of Economic Studies*.

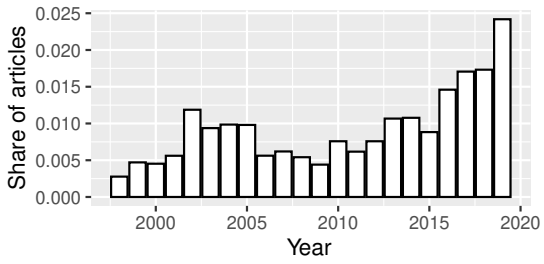
3.4 Constant but large share of top publications by top-five authors

We shed light on the association of affiliations to top-five universities and publications in top journals, defined as a subset of the top-five journals: *Quarterly Journal of Economics, Journal of Political Economy* and *Review of Economic Studies*. The other two top-five journals do not provide their articles in HTML (cf. Section 2). On average, around a third of all empirical articles in the top-five journals have at least one affiliation associated with a top-five university in the respective publication year (Figure 4). This share remained stable over the last two decades, hinting at a certain amount of rigidity in the publication system of economics. Our finding is in the same ballpark as previous literature (e.g., Heckman and Moktan 2020; Kocher and Sutter 2001).

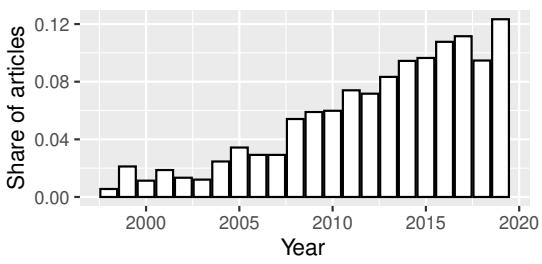
4 Credibility revolution

4.1 Increasing focus on causality

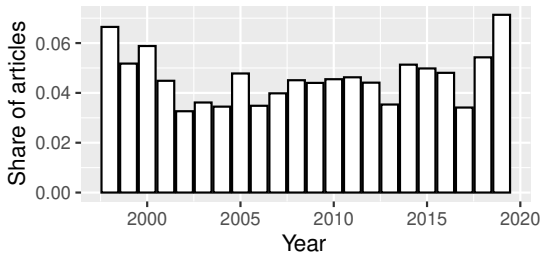
Selecting the appropriate methods is crucial to the credibility of a study. In recent years, we observe a greater emphasis on experimental and quasi-experimental methods (cf. Figure 5). About half of the empirical articles in 2019 mention the identification of a (causal) effect and controlling for unobserved heterogeneity using fixed effects. The rate of increase varies among the quasi-experimental methods. While the share of articles mentioning difference-in-differences or regression discontinuity grows steadily, the share of articles mentioning instrumental variables grows less pronounced but reaches almost 30 % in 2019. However, this approach might be more prone to *p*-hacking than other methods (Brodeur et al. 2020). Moreover, there is an increasing trend in the use of experimental methods, e.g., lab or field experiments. This might be attributed to public agencies and private companies being more willing to fund such experiments (Stock 2010). However, these experiments are still rare, with only 6 % of all articles mentioning a field experiment



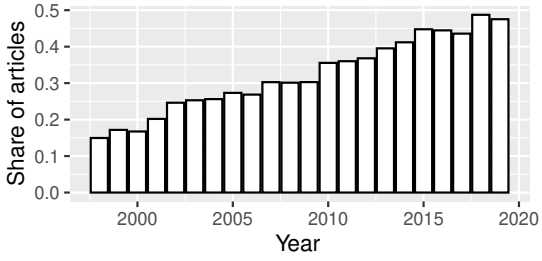
(a) Bunching



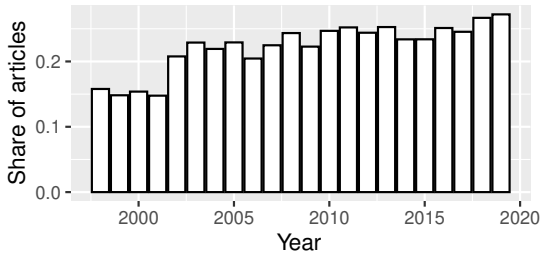
(b) Difference-in-differences



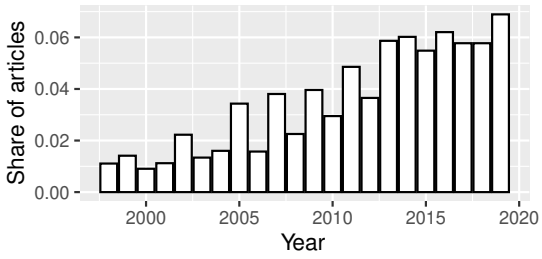
(c) Event study



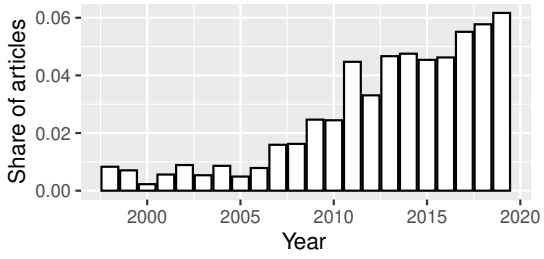
(d) Identification



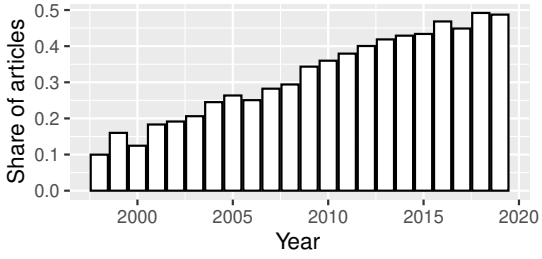
(e) Instrumental variables



(f) Lab experiment

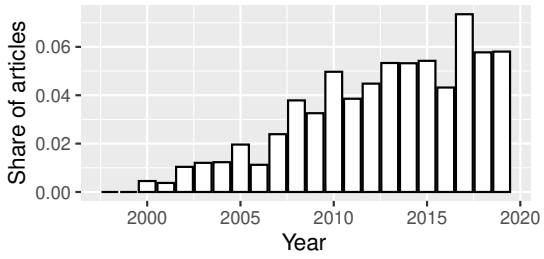


(g) Field experiment

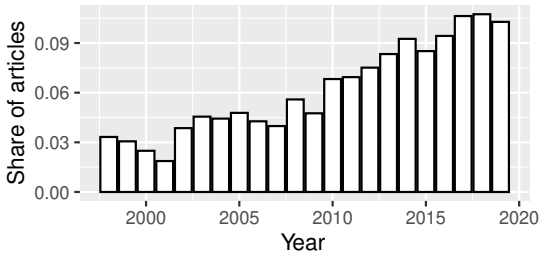


(h) Fixed effects

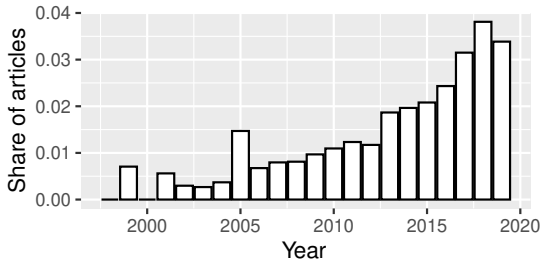
and another 6 % mentioning a lab experiment. The rise in the regression discontinuity design might be attributable to the influential article by Angrist and Lavy (1999) that analyzes the relationship between class sizes and test scores. Synthetic control methods are still relatively new to economics, aiming to fulfill the parallel trend assumption in the difference-in-differences design by creating an additional synthetic unit using linear combinations of non-treated and not-yet treated units. The modest increase in synthetic control methods might be credited to Abadie et al. (2010), who applied it to estimate the effect of the California Tobacco Control Program.[11] In line with the previous literature,



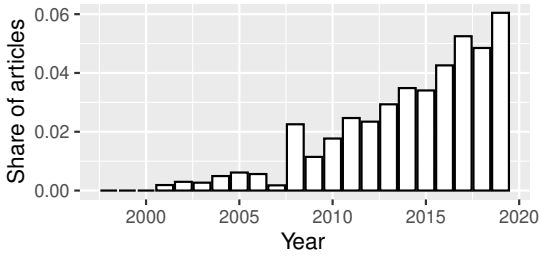
(i) Matching



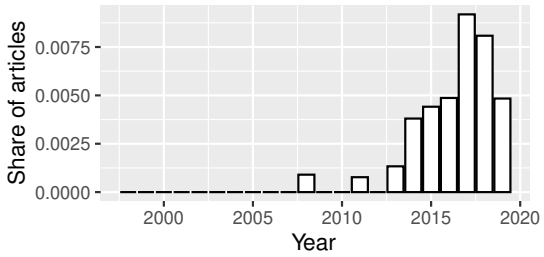
(j) Quasi- and natural experiments



(k) Randomized controlled trial



(l) Regression discontinuity



(m) Synthetic control method

Figure 5: Word trends related to causality analysis (empirical articles: 22k)

Notes: Figures show shares of empirical articles mentioning (a) bunching, (b) difference-in-differences, (c) event study, (d) identification, (e) instrumental variables, (f) lab experiment, (g) field experiment, (h) fixed effects, (i) matching, (j) quasi- and natural experiments, (k) randomized controlled trial, (l) regression discontinuity, and (m) synthetic control method. Full regular expressions can be found in Table B.6 in the Online Appendix.

we see a decline in the use of synthetic control methods after the initial increase that began in 2010 (Goldsmith-Pinkham 2024).

In sum, there has been a growing focus on causality in the last two decades in economics. Field experiments are steadily increasing. The identification of causal effects by using quasi-experimental designs is constantly increasing over time, as well. Our finding is in line with previous literature that focused on working papers (Garg and Fetzer 2025; Goldsmith-Pinkham 2024; Currie et al. 2020).

4.2 Improvement of econometric methods

One key driver of the credibility revolution is the steady advancement of econometric methods, making econometric research more robust to non-normally distributed or clus-

[11] Abadie and Gardeazabal (2003) introduced the synthetic control methods a few years earlier.

tered data. This includes the usage of clustered inference, fixed effects, as well as semi- and non-parametric methods.[12] It also includes correcting for multiple hypothesis testing to improve the credibility of estimation results, especially as the number of statistical tests per article increases (cf. Figure 7b).

We see a notable increase in articles reporting tests with clustered inference (cf. Figure 6a). By 2019, more than 40 % of the articles fall into this category. Clustered inference is used when the data have a hierarchical structure or when observations within clusters are likely to be correlated. This approach helps to prevent biased standard errors due to intra-cluster correlation, ensuring more reliable statistical inferences. The rise in the usage of clustered inference might be related to the study by Bertrand et al. (2004), which shows that ordinary least squares regression with clustered standard errors considerably reduces sampling uncertainty. Our regular expressions include different versions of 'cluster' which are likely to capture clustered inference more generally, including clustered standard errors and confidence intervals. The rise in clustered inference is also reflected in bootstrapping methods being mentioned more frequently over time (cf. Figure C.2).

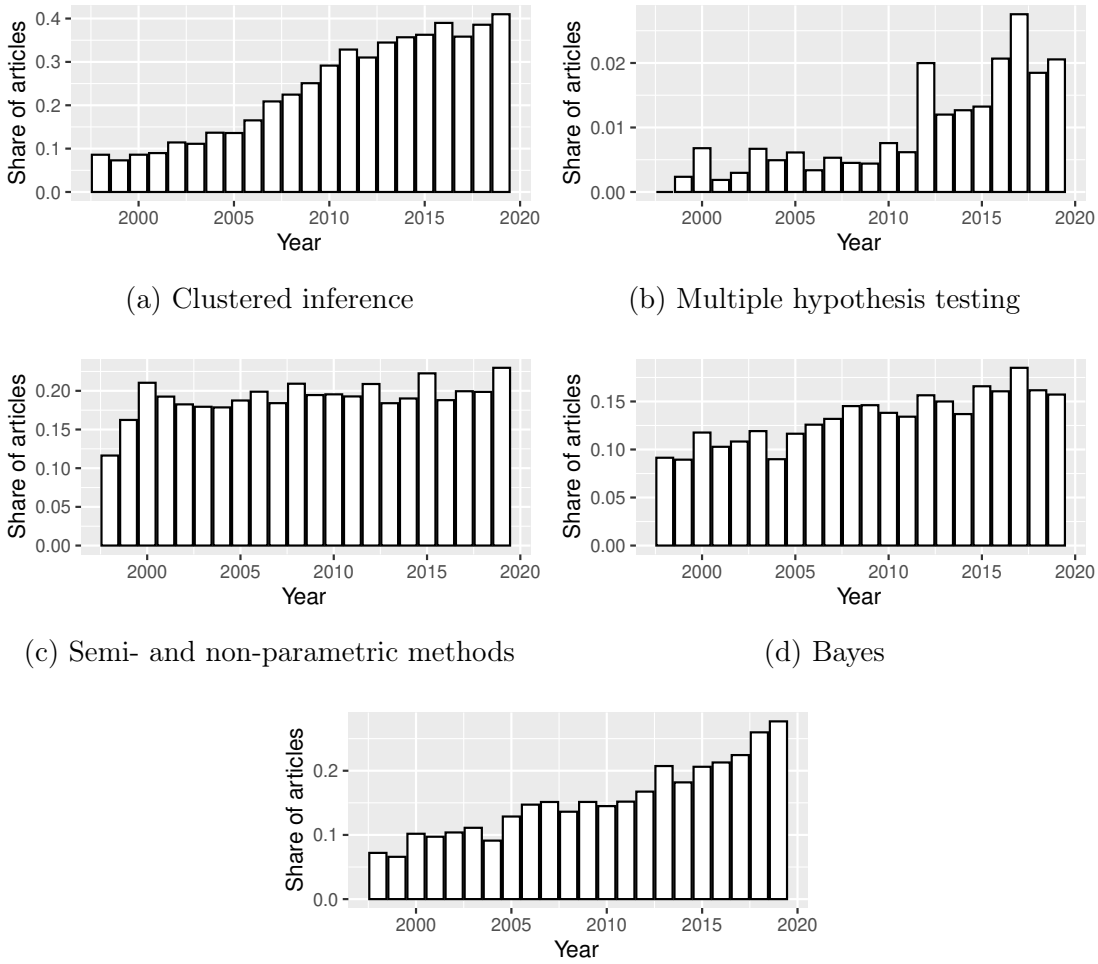
As mentioned above, there has also been a large increase in the usage of fixed effects (cf. Figure 5h). From 1998 to 2019, the number of empirical articles mentioning fixed effects quintupled, culminating in nearly half of the articles mentioning fixed effects.

When multiple tests are conducted, the probability of falsely rejecting at least one true null hypothesis increases. To address this, various correction methods, such as the Bonferroni correction (Dunn 1961) or the False Discovery Rate approach (Benjamini and Hochberg 1995), are used to adjust significance levels. This ensures that the overall error rate is kept within acceptable limits, thereby providing more reliable and valid results. However, correcting for multiple hypothesis testing is still uncommon in economics and other disciplines. In 2019, only around 2 % of all articles mention multiple hypothesis testing (cf. Figure 6b). In comparison, 0.33 % of all psychology articles mention any type of correction (Nuijten et al. 2016). The trend in economics is increasing, with a spike after 2010. This spike might be due to Romano et al. (2010), who provided a comprehensive overview of current methods for multiple hypothesis correction and highlighted unresolved issues.

Semi- and non-parametric methods, which make fewer assumptions about the underlying data distribution, offer more flexible and robust analytical tools, enhancing the credibility of empirical research. These methods are well-established in economics. Around 20 % of all empirical articles mention semi- or non-parametric methods (cf. Figure 6c). The trend remains steady, indicating that these methods have long been integral to economics research.

We also observe an increased mention of Bayesian methods (cf. Figure 6d) and confidence intervals (cf. Figure 6e). By 2019, over 15 % of all empirical articles mention Bayes, and more than 25 % mention confidence intervals. This can be seen as a shift toward more advanced methods during the credibility revolution, but also toward a larger focus on effect sizes. Compared to Poirier (2006), we find an increase not only in theoretical and mathematical Bayesian articles but also in empirical Bayesian articles. Given their more common use in medicine, the surge in Bayesian methods might be attributed to health economists as early adopters (e.g., Briggs 1999). Moreover, the increased focus on confidence intervals, which emphasize effect size over statistical significance, indicates a growing attention to effect sizes (cf. Section 5.2).

[12] Our regular expression for non-parametric methods searches for the expression 'non-parametric'. Hence, modern machine learning techniques are not explicitly included.



(a) Clustered inference

(b) Multiple hypothesis testing

(c) Semi- and non-parametric methods

(d) Bayes

(e) Confidence intervals

Figure 6: Word trends related to econometric methods (empirical articles: 22k)

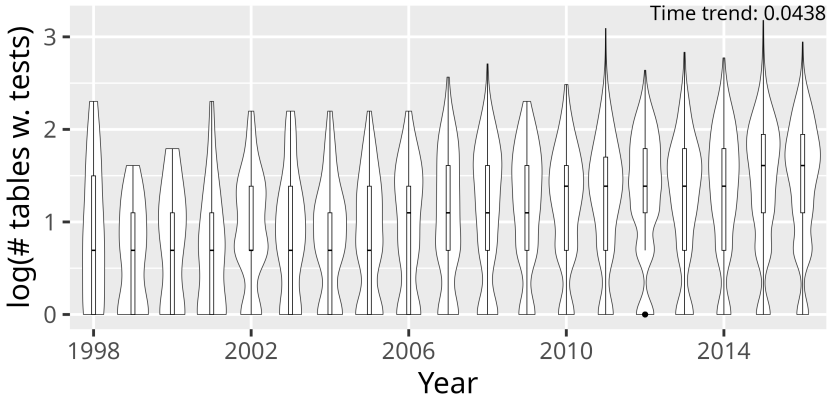
Notes: Figures show share of empirical articles mentioning (a) clustered inference, (b) multiple hypothesis testing, (c) semi- and non-parametric methods, (d) Bayes, and (e) confidence intervals. Full regular expressions can be found in Table B.6 in the Online Appendix.

4.3 Increasing number of tables and statistical tests

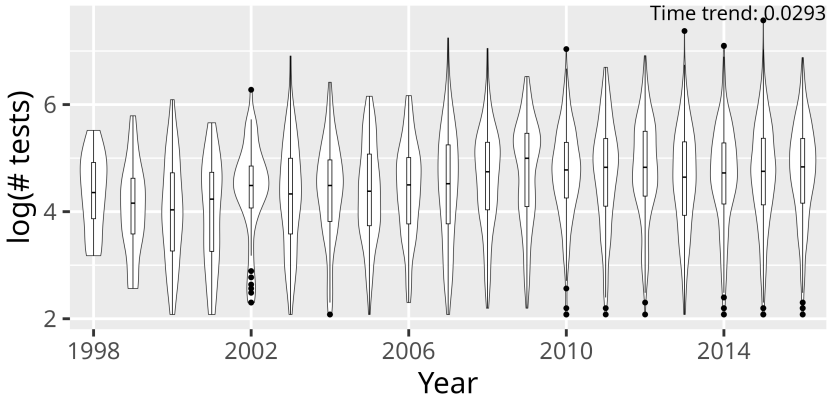
From 1998 to 2016, there is a noticeable increase in the number of tables with statistical tests per article (cf. Figure 7a). While the median article in 1998 had two tables with tests, the median article in 2016 had five tables with tests. Similarly, the number of tests per article also exhibited an upward trend (cf. Figure 7b). In 2000, the median article comprises 78 tests. In 2016, the number rises to 126. These additional tables and statistical tests may represent an increasing number of robustness checks and sensitivity analyses.

Notably, the number of pages per article does not increase over time (cf. Figure 7c). In fact, there appears to be a slight decrease, even after accounting for variations between journals. This decrease is driven primarily by non-top-five journals in our sample. Previous literature has noted an increase in pages for the top-five journals (Card and DellaVigna 2013); our sample confirms this for journals such as the *Quarterly Journal of Economics*, *Journal of Political Economy*, and *Review of Economic Studies*. Given the rising number

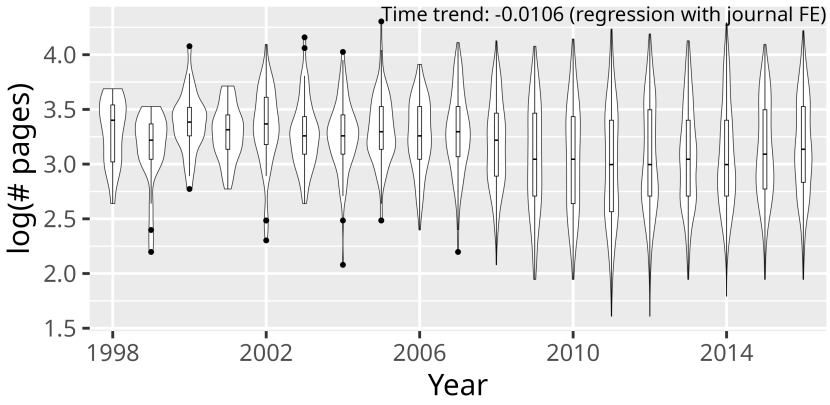
of tests and the increasing word count in full texts (Bruns et al. 2024b), it can be assumed that the content per page has increased over time, especially in non-top-five journals.



(a) Number of tables with tests per article



(b) Number of tests per article



(c) Number of pages per article

Figure 7: Article length and the number of tables and statistical tests (empirical articles with statistical information: 4k)

Notes: Figures show violin plots related to (a) the number of tables with tests, (b) the number of tests, and (c) the number of pages per article. The time trend shown in Figure (c) uses journal fixed effects. Without journal fixed effects the time trend is: -0.0127.

5 Open science movement

5.1 Slightly increasing awareness of the replication crisis

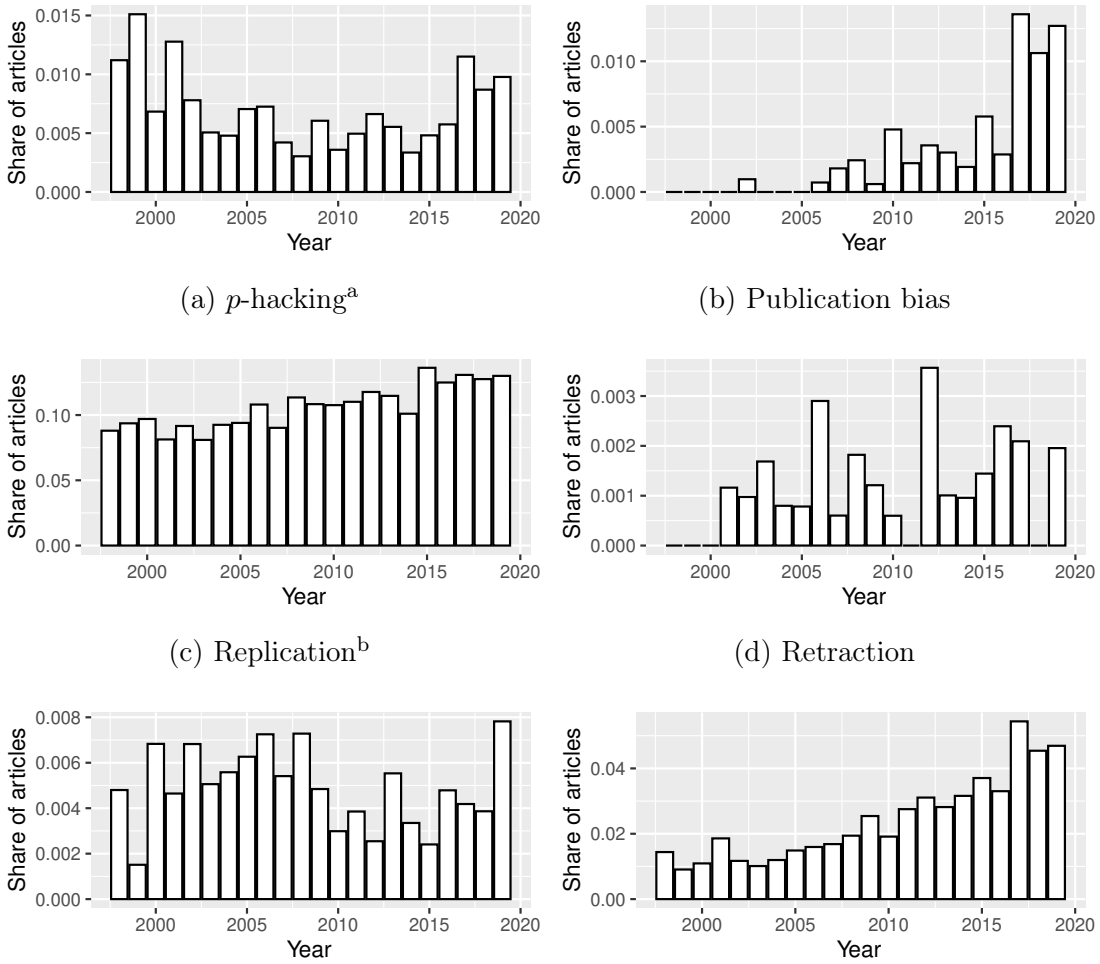
One of the first replication studies was carried out by Dewald et al. (1986), providing initial evidence of low replication rates in economics. Since the starting point of our sample in 1998, there has been a slight increase in awareness of the replication crisis, as indicated by the growing share of articles discussing replications with a prevalence of about 13 % at the end of the sample (cf. Figure 8). Additionally, biases in the research and publication process have garnered increased attention, observable in the trends for terms such as 'publication bias' and 'specification searching'. However, reproducibility does not appear to be a highly relevant topic in economics articles. The movement toward open science seems to be driven mostly on the journal-side, as can be seen by the increasing number of data and code availability policies (cf. Section 5.4). We also observe a rising trend in the proportion of articles employing the term 'statistical power', reflecting a more introspective stance on data analysis quality. This trend may be correlated with several influential economics papers that collectively argue that the discipline is underpowered (Ioannidis et al. 2017; Ziliak and McCloskey 2004; McCloskey and Ziliak 1996). Conversely, retractions are rarely discussed within the economics literature, and no distinct trend is observed (Ankel-Peters et al. 2025). While our findings suggest only a slight increase in the awareness of the replication crisis, a survey among authors publishing in top economics journals as well as PhD students from the top 20 North American Economics departments suggests that transparency and reproducibility practices may be on the rise in economics (Swanson et al. 2020; Miguel 2021). This discrepancy might be explained by the survey focusing at the top of economics research, while our data provides a broader and more inclusive assessment.

5.2 Larger sample sizes and smaller effect sizes

It is crucial to differentiate between statistical and economic significance. Policies based on results that are statistically significant but have very small effect sizes may prove ineffective (McCloskey and Ziliak 1996). As effect sizes in economics are often not standardized, it is difficult to measure whether effect sizes are actually discussed in the article. The share of articles explicitly mentioning effect sizes is low and increases from 1 % in 2005 to 7 % in 2019, while around 30 % of all empirical articles mention elasticities over the time span of our sample. These findings tend to be consistent with previous research, which found no increase in the distinction between economic and statistical significance in the *American Economic Review* between 1996 and 2004, and a generally low prevalence of discussions about effect sizes (Ziliak and McCloskey 2004).

DORIS is able to extract the maximum number of observations given in a table. Among all empirical articles with statistical information, we observe an annual increase in the maximum number of observations per article of around 0.16 on the log scale (cf. Figure 9). This trend could be related to the growing accessibility of data, particularly big data.

At the same time, we observe a slight decrease in statistical significance of the main tests per empirical article with statistical information despite generally increasing sample sizes (cf. Figure 10). Per year, we measure an average decrease in the z-values of around 0.006. After translating the z-values to typical p-value threshold categories, we see that the number of main tests falling under a certain p-value threshold remains stable over time



(a) p-hacking[a] (b) Publication bias

(c) Replication[b] (d) Retraction

(e) Specification searching (f) Statistical power

Figure 8: Word trends related to the replication crisis (all articles: 31k)

Notes: Figures show share of articles mentioning (a) p-hacking, (b) publication bias, (c) replication, (d) retraction, (e) specification searching, and (f) statistical power. Full regular expressions can be found in Table B.6 in the Online Appendix. a: Includes p-hacking, data dredging, data snooping, data fiddling, selective reporting, and significance chasing. b: Includes replication and reproducibility.

(cf. Figure 11). This suggests that, although there is a slight decline in significance within the categories of statistical significance, the distribution of tests across these categories does not change. With increasing sample sizes, we would actually expect p-values to become smaller. This might be explained by decreasing effect sizes, as discussed in the next paragraph.

We find a slight decrease in the effect sizes of the main tests per year (cf. Figure 12). We measure the effect size using the partial correlation coefficient (PCC). The PCC is calculated as follows:

$$\text{PCC} = \frac{t}{\sqrt{t^2 + df}}, \tag{1}$$

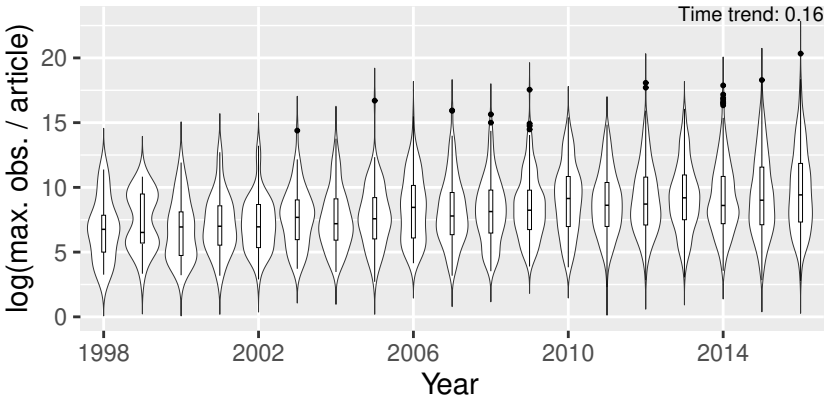


Figure 9: Number of observations (empirical articles with statistical information: 4k)

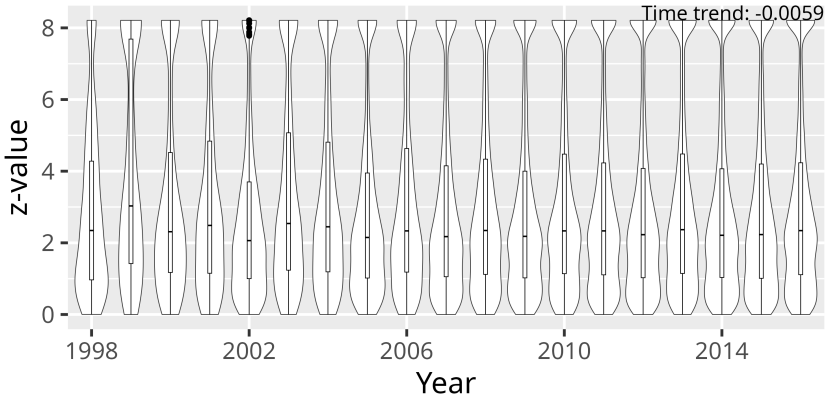


Figure 10: z-values of main tests (empirical articles with statistical information: 4k)

where df are the degrees of freedom and t is the t-value of the respective test. More information on the PCC can be found in the meta-research literature (Doucouliagos 2011). The average annual decrease in the PCC is around 0.07 on the log scale.[13]

Due to publication and reporting bias, half of the effect sizes in economics are typically inflated by a factor of two or more (Ioannidis et al. 2017). One mechanism is that researchers may run multiple experiments or empirical specifications based on relatively small samples and selectively report those that generate large and statistically significant effects by chance. Because effect estimates from small samples are inherently more variable, this process can lead to systematically exaggerated published effect sizes. We find indications that researchers publish smaller effect sizes with increasing sample sizes over time. Since larger samples should lead to smaller p-values *ceteris paribus*, increasing accessibility to big data might have helped reduce the exaggeration of effect sizes due to p-hacking, potentially making reported effect sizes in economics more reliable. Many of the published effect sizes, which are potentially exaggerated to achieve statistical significance, might be economically non-significant (Ziliak and McCloskey 2004).

[13]We use the log scale for better visualization. Since the PCC ranges between 0 and 1, values would accumulate near 0 without the log scale.

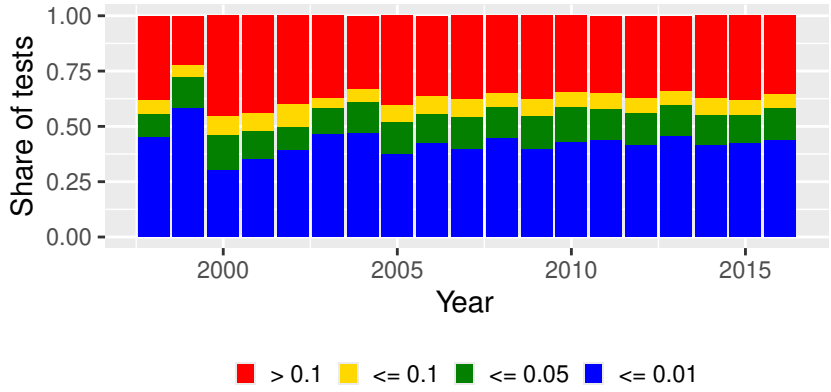


Figure 11: Number of main tests falling under a certain p-value threshold (empirical articles with statistical information: 4k)

Notes: Only tests reported with standard significance levels (10, 5, and 1 percent) are considered.

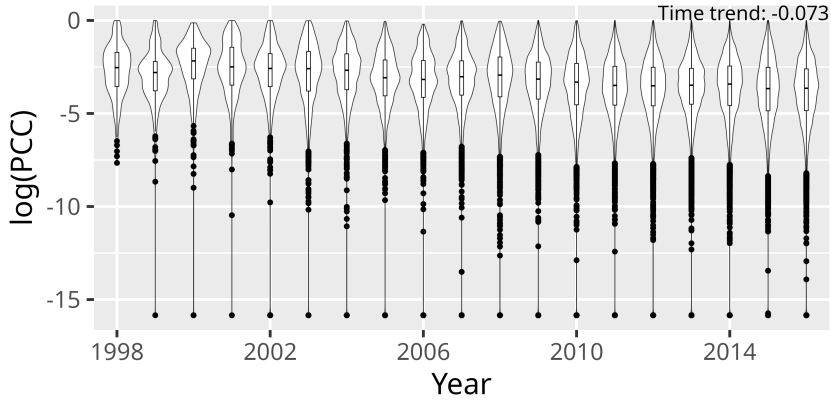


Figure 12: Partial correlation coefficients (PCC) of main tests (empirical articles with statistical information: 4k)

5.3 More null findings get published

Economics research tends to favor significant results. This phenomenon is known as publication bias (e.g., Bruns et al. 2024a; Brodeur et al. 2023; Pütz and Bruns 2021; Blanco-Perez and Brodeur 2020; Andrews and Kasy 2019; Stanley 2005). Researchers either keep their non-significant studies in their file drawer or these studies are rejected during the review process (Rosenthal 1979). To achieve full transparency in research, it is important that non-significant results, or null results, are also available to the community if they are based on meaningful research questions and adequate research designs. We observe a steep increase in mentioning null findings, but the prevalence is still rare with around 2 % in 2019 (cf. Figure 13). This rise may be related to the year 2015, when economics journals began to issue editorial statements aimed at reducing the penalty on null findings (Chopra et al. 2023) and the extent of specification searching (Blanco-Perez and Brodeur 2020). As these statements might have already been discussed at conferences before, there could be an anticipation effect, explaining the relatively high prevalence from 2010 to 2014. Our finding obtained for a broad range of economics journals is in contrast

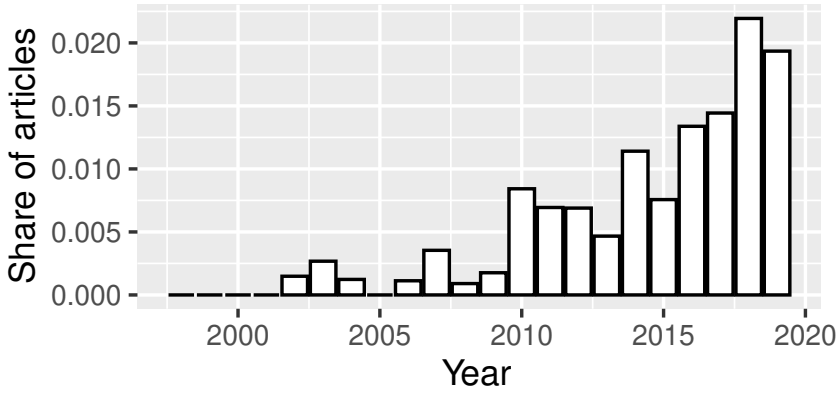


Figure 13: Share of empirical articles mentioning null findings (empirical articles: 22k)

to another study in economics (Garg and Fetzer 2025) which shows a decline of null results over the last decades, using working papers from two databases (CEPR and NBER).

5.4 Open data and code

5.4.1 Data and/or code availability policies are increasing

In times of the replication crisis, sharing data and code becomes especially crucial. Journal policies on mandatory data and code availability require authors to make their data and code accessible. Facilitating replication efforts within the research community can help curb dubious research methods, amend findings that have already been published (e.g., Bach et al. 2023; Matray and Boissel 2023), and broadly enhance credibility (e.g., Brodeur et al. 2024a; Askarov et al. 2023).

With the prospect of replication, researchers tend to avoid questionable research practices and are more inclined to report null results. Data and code availability policies might also enhance the receptivity of null results among editors and reviewers (Brodeur et al. 2024b; Askarov et al. 2023). Concurrently, researchers are encouraged to thoroughly verify their findings to prevent errors and to carry out additional robustness checks, as other scholars can easily identify errors and weak conclusions through access to shared data and code (Askarov et al. 2023).

As of 2021, only about 20 % of all empirically-oriented economic journals have a mandatory data availability policy (Vlaeminck 2021). In our sample, almost one half of the journals implemented a data and/or code availability policy (cf. Figure 14). It seems that the journals from the top 50 considered here are more inclined to implement these policies. The starting point for leading journals is marked by the *American Economic Review* and *Econometrica* announcing their mandatory data- and code-sharing policy in 2004. The demand for these policies is imminent as there is a steady increase in web requests to data archives, e.g., the data archive of the *Journal of Applied Econometrics*.[14]

Additionally, the *Journal of Applied Econometrics* implemented open science badges in 2018 as the only economics journal to date. These badges are usually used within natural sciences and psychology as an incentive and highlight if a study either provided open data, open code, or was pre-registered. A recent study found that open science

[14]Cf. http://qed.econ.queensu.ca/jae/summary-stats.html, last retrieved on 31.08.2024.

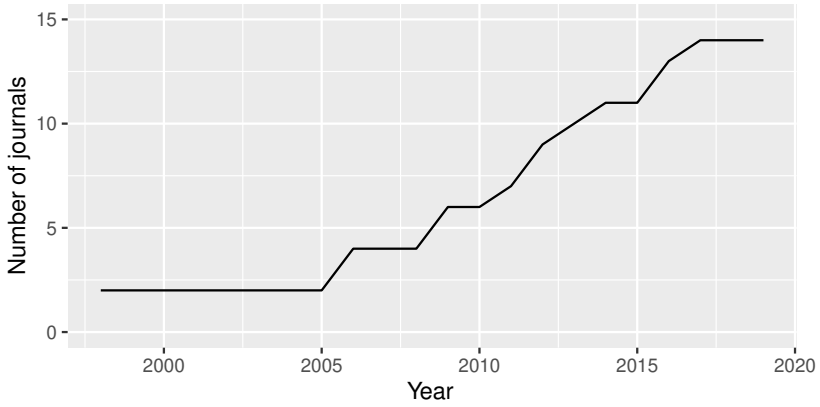


Figure 14: Mandatory data and/or code availability policies (empirical articles: 31k)

Table 2: Data and/or code availability over time in percentages (matched articles with data and/or code: 1.5k)

	2011	2012	2013	2014	2015	2016	2017	2018	2019
Data and code not shared	84.62	79.45	76.71	68.83	38.82	35.33	31.43	23.26	14.95
Data and code shared	15.38	20.55	23.29	31.17	61.18	64.67	68.57	76.74	85.05

Note: As identified by the software *Find Economic Articles with Data*, cf. Section 2.1 for details.

badges significantly influenced social media attention, but they did not notably affect the research impact (Zong et al. 2023).

5.4.2 Data and/or code availability policies are effective

Analyzing all matched articles with data and/or code, we find that data and/or code availability policies are highly effective in creating open data and/or code. Although less than 1 % of all articles shared data and/or code in the absence of an active data and/or code availability policy, around 60 % of the articles shared data and/or code when such a policy was in place. This highlights the long-term effectiveness of these policies, evidenced by the fact that in 2019 even 85 % of all matched empirical articles shared data and/or code publicly (cf. Table 2). However, there are various exceptions to these policies, e.g., when using proprietary data. In addition, most policies are not uniform, ranging from mere encouragement to strict enforcement. To achieve even greater coverage, data and code availability policies need to be strictly enforced and unified (Bruns et al. 2023; Askarov et al. 2023; Christensen et al. 2019). In a random sample of 250 articles each, analysis scripts were available for 1 % and raw data for 7 % in the social sciences (Hardwicke et al. 2020), and for 1 % and 2 % in psychology, respectively (Hardwicke et al. 2022). These numbers align with those found in economics for journals without a data and code availability policy.

As additional evidence for the larger data sets mentioned in the previous section, we find a steady increase in the file sizes of the replication packages provided by the authors of articles with shared data and code (cf. Figure 15). While the median file size in 2005 was 0.32 MB, in 2023 it was already 21.84 MB.

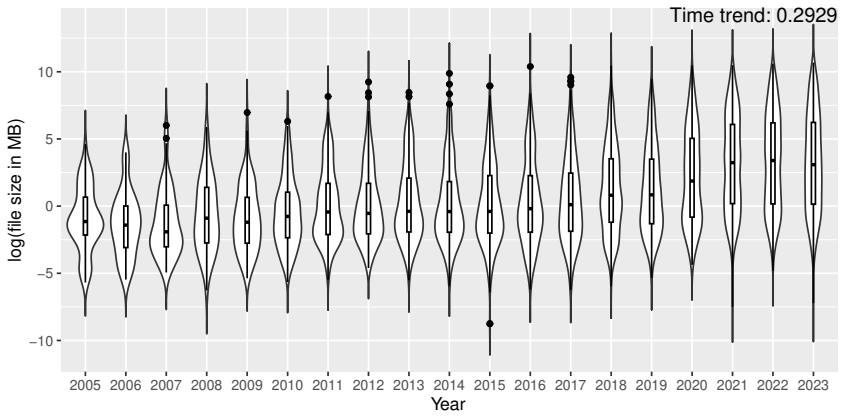
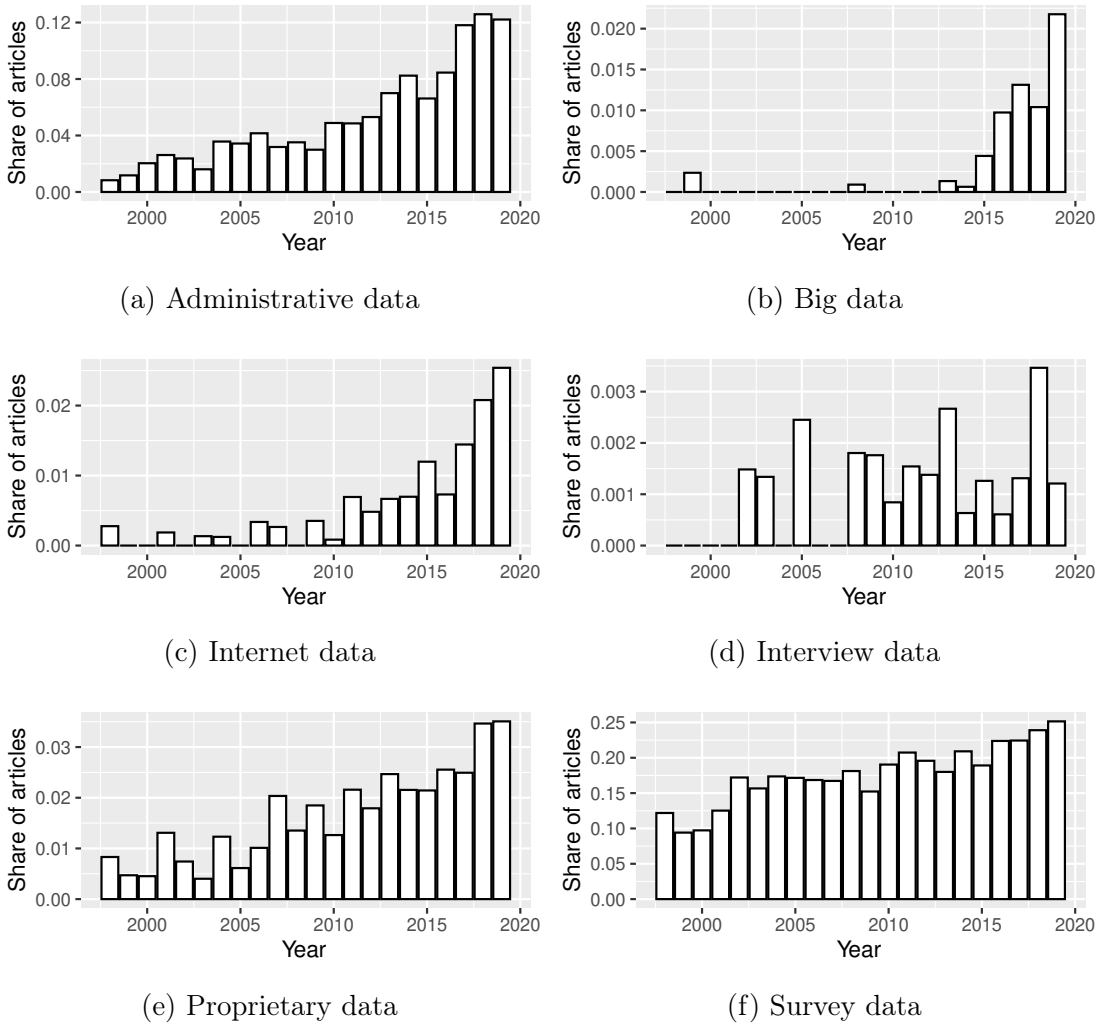


Figure 15: Disc space of shared data and/or code (articles with data and/or code: 11k)

5.5 Increased use of proprietary data is a challenge to open science

As articles become more empirical over time, we see an increase in the variety of data sets used (cf. Figure 16). The most commonly used type of data is survey data, with a prevalence of around 10 % before the year 2000 and more than 25 % in 2019. The usage of administrative data grew from below 5 % to almost 12.5 %. Open science depends on data transparency. Since administrative data are often proprietary, their increased usage challenges open science. Previous literature also found a decreasing share of empirical work based on publicly available data sets among three of the top-five journals (Hamermesh 2013). As a result, there are growing calls to expand access to administrative data for researchers by ensuring conditions such as direct access to de-identified microdata through local statistical offices (Card et al. 2011). In contrast, administrative data, as well as the increased usage of internet data, allow for larger sample sizes and, thus, for more robust analyses. This is also reflected in an increasing tendency to use the term 'big data'. Interview data have hardly been used in the past two decades. Our findings, especially regarding proprietary data, align with recent studies that focus on fewer journals than our study or on working papers (Garg and Fetzer 2025; Currie et al. 2020). In summary, we see a shift toward larger or big data, but also to more proprietary data that might bypass current data availability policies implemented by journals, as proprietary data often permit exceptions.



(a) Administrative data (b) Big data

(c) Internet data (d) Interview data

(e) Proprietary data (f) Survey data

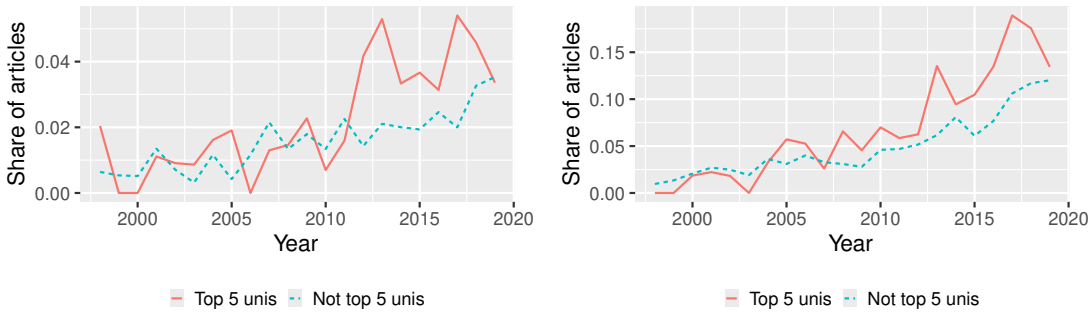
Figure 16: Data types used in empirical articles (empirical articles: 22k)

Notes: Figures show share of empirical articles mentioning (a) administrative data, (b) big data, (c) internet data, (d) interview data, (e) proprietary data, and (f) survey data. Full regular expressions can be found in Table B.6 in the Online Appendix.

We also analyze the data usage in empirical articles with at least one affiliation to a top-five university in the respective publication year (Figure 17). While administrative data are more commonly used than proprietary data, we see a growing difference between articles with at least one affiliation to a top-five university and articles without such an affiliation since 2010. As administrative as well as proprietary data are often linked to additional costs, this finding is opposed to the pillars of open science, indicating signs of an unlevel playing field in economics.

5.6 Almost no distinction between hypothesis-testing and exploratory research in economics

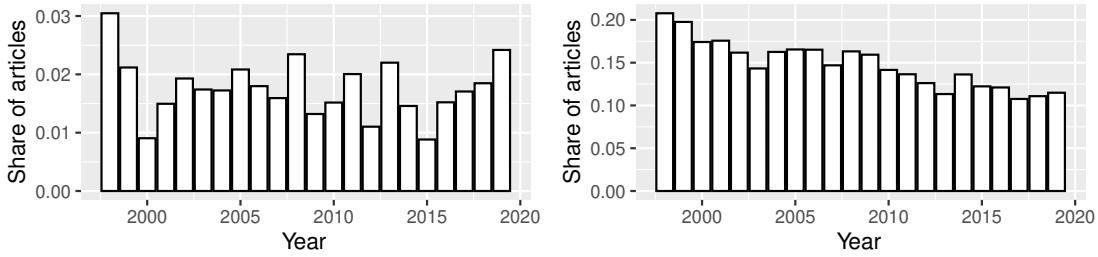
Hypothesizing after the results are known (also known as HARKing, Kerr 1998) is a potentially questionable research practice (Munafò et al. 2017). For example, a researcher could propose a hypothesis about the correlation between two factors after accidentally



(a) Proprietary data (b) Administrative data

Figure 17: Usage of data types distinguished by top-five and not top-five universities (empirical articles: 22k)

Notes: Figures show share of empirical articles mentioning the usage of (a) proprietary data and (b) administrative data distinguished by top-five and not top-five universities in the specific year. Full regular expressions can be found in Table B.6 in the Online Appendix.



(a) Exploratory (b) Hypothesis testing

Figure 18: Distinction between exploratory and hypothesis-testing research in empirical articles (empirical articles: 22k)

Notes: Figures show share of empirical articles mentioning (a) exploratory research and (b) hypothesis-testing research. Full regular expressions can be found in Table B.6 in the Online Appendix.

finding such a correlation while analyzing a different research question. Framing this empirical finding as hypothesis testing inflates the underlying type I error (Kerr 1998). A clear distinction between exploratory research and hypothesis-testing research can help prevent HARKing. Only a small share of the empirical articles states that their research is exploratory. There is no trend over time, and the share rarely exceeds 2 % (cf. Figure 18). In contrast, a larger share of empirical articles declares their research as hypothesis-testing research. However, the trend over time is negative, decreasing from around 20 % in 1998 to around 10 % in 2019.

Pre-analysis plans can help prevent framing exploratory research as hypothesis-testing research (Imbens 2021; Duflo et al. 2020; Olken 2015). By declaring hypotheses and methods publicly before analyzing the data, it is possible to reduce questionable research practices, such as *p*-hacking, specification searching, and HARKing. However, pre-analysis plans are still uncommon in economics. The first occurrence of pre-analysis plans in our sample is in 2012 with a growing prevalence from 2016 onward (cf. Figure C.1 in the Online Appendix). Still, in 2019 not even 0.25 % of all empirical articles mention any pre-analysis plans. Overall, it seems that the debate between exploratory hypothesis

testing and confirmatory hypothesis testing, as well as a culture of pre-analysis plans, is more common in other disciplines (e.g., Rubin and Donkin (2022) for psychology or Ferguson et al. (2023) for an overview about all social sciences). Within economics, pre-analysis plans are more prevalent in the context of randomized controlled trials (Brodeur et al. 2024b).

5.7 Slow movement to open source software, but faster movement to open access

In empirical articles or simulation studies, programming languages or statistical software are used for data analysis. We find that only around 16 % of the empirical articles mention the software they used in their full text (cf. Figure 19). However, the software plays a crucial role in data analysis, as findings could differ due to precision differences or different default methods. Among the articles that mention their software, Stata and MATLAB are the most important statistical software, both being proprietary. We see a notable upward trend regarding the open source programming language R, while the popularity of the proprietary programming language GAUSS vanishes. This suggests that free open source software becomes steadily more popular in economics while still being dominated by proprietary software that might hinder reproducibility. Open source software is a crucial part of open science (UNESCO 2021), as it allows researchers to examine the underlying code, potentially reducing the likelihood of computational errors and thus enhancing replicability.

The share of publicly available articles is increasing and spikes at around 35-40 % at the end of the sample period (cf. Figure 20). This is similar to what has been found in the social sciences (40 %, Hardwicke et al. (2020)) but lower than in psychology (65 %, Hardwicke et al. (2022)). We measure public availability as indicated by the Web of Science. This includes the availability of submitted or accepted versions of the articles at institutional repositories.

6 Conclusions

We map the development of research in economics by analyzing 578,132 statistical tests from 3,746 articles and the full text of 30,675 articles from the top 50 economics journals from 1998 to 2019. We connect scientometrics with individual statistical data to take stock of the state of economics research and to identify areas for further improvement.

General trends

Although economics has become more empirical, the range of research topics has largely remained stable. The increasing number of co-authors might be explained by the advent of the internet and the availability of affordable travel options that have lowered the costs related to working with colleagues from distant locations (Hamermesh 2013). The rise of big data could have also increased the demand for non-economist specialists, such as data engineers or data scientists, who may be granted co-authorship.

Diversity within teams is increasing, as indicated by an increasing number of articles with female authors. Yet, there is a large room for improvement given the importance of a diverse author team, as, for instance, articles tend to be more readable with an increasing share of female authors (Hengel 2022) and also more cited (Bruns et al. 2025).

The share of articles in three top journals (*Quarterly Journal of Economics, Journal of Political Economy*, and *Review of Economic Studies*) with at least one affiliation to a

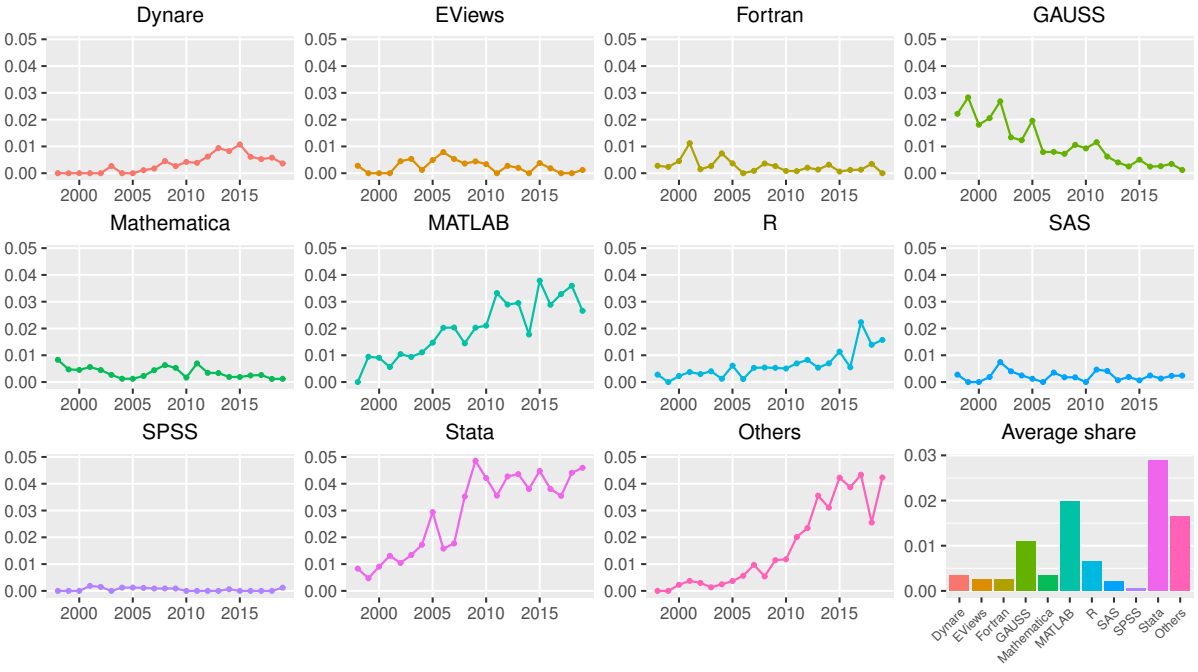


Figure 19: Share of empirical articles using specific software (empirical articles: 22k, only 3,563 articles that mention software)

top-five university is large (30 - 40 %) and relatively constant over time. Our findings are in line with previous research that promotes the image of the top-five journals as closed clubs (e.g., Lutmar and Reingewertz 2021; Winkler et al. 2014; Cardoso et al. 2010; Rupp and Jr. 2002; Scott and Mitias 1996).

Credibility revolution

The credibility revolution focuses mainly on the use of better methods to detect causal relationships. We find that quasi-experimental designs, such as difference-in-differences designs or regression discontinuity designs, and experimental designs, both in the lab and field, have gained importance. However, this increased focus on experimental and

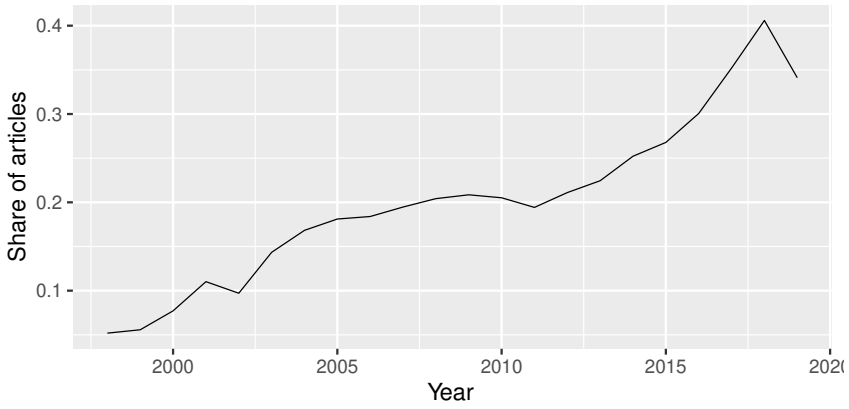


Figure 20: Share of publicly available articles (all articles: 31k, but due to incomplete matching with Web of Science: 28,649 articles)

quasi-experimental methods may come at a price. By choosing the method prior to the topic, research might have narrowed down to less important or fewer topics, potentially overshadowing high-quality descriptive analyses that address complex real-world issues (Gopalan et al. 2020).

Our analysis also indicates that econometric methods are improving. The usage of clustered inference, Bayesian techniques, and non-parametric approaches increases over time. Another beneficial development is the increasing number of tables and statistical tests being reported within an article, pointing to more robustness and sensitivity analyses. Large room for improvement remains for the correction of multiple hypothesis testing.

Previous literature showed that economics articles become longer in the top-five journals (Card and DellaVigna 2013), but longer articles can be more difficult for readers to digest and take longer to review (Imbens et al. 2019). Our analysis reveals that the length of articles that are not published in the top-five is constant or even decreasing over time.

Open science movement

Awareness of the replication crisis and the need for open science seems to increase only slightly in economics, as indicated by the use of expressions such as *p*-hacking, statistical power, or replication. The main agents in advancing open science and addressing low replication rates seem to be journal editors by introducing data and code availability policies and statements on the merit of null findings (Blanco-Perez and Brodeur 2020). While our analysis shows that data and code availability policies are increasing, only about 20 % of empirically oriented economic journals have a mandatory data policy (Vlaeminck 2021). Low availability of data and code is also common in the social sciences and psychology (Hardwicke et al. 2020; Hardwicke et al. 2022). This leaves room for further improvement to ensure reproducibility of published research.

Publishing null results is uncommon in economics. While it is desirable to increase the number of null results being published, they should be accompanied by an a priori power analysis to demonstrate that the null finding might be interpreted as the absence of an effect (Bruns et al. 2024a).

While open code policies have improved code availability, code quality should not be neglected. Reproducible code must not only be well-commented, an instruction manual is also needed. 'The Turing Way' is a useful guide to learn tried and tested techniques for more reproducible science.[15] Virtual environments such as *renv* are advised to diminish replication differences due to different software or package versions.[16] Since results might even differ depending on the operating system and the exact machine, virtual containers such as *Docker* should be considered as well.[17] With regard to data availability, data management should be embedded in the FAIR framework, making the data findable, accessible, interoperable, and reusable (Wilkinson et al. 2016).

A special challenge to open data is proprietary data that cannot be shared in replication packages. Potential solutions include de-identified data that could be shared (Card et al. 2011). Of particular concern is that authors with top-five affiliations are increasingly using proprietary data, which might potentially unlevel the playing field in economics.

Compared to other disciplines, economics does not try to obtain credibility and transparency by implementing pre-analysis plans. They might be perceived as too restrictive. For example, in observational research, it could be difficult to create an adequate model without first knowing the data. A promising approach could be analysis blinding

[15] Cf. https://book.the-turing-way.org/, last retrieved 31.08.2024.
[16] Cf. https://rstudio.github.io/renv/articles/renv.html, last retrieved 31.08.2024.
[17] Cf. https://www.docker.com/, last retrieved 31.08.2024.

(Sarafoglou et al. 2023). As a first step, the statistical analysis is developed and finalized on data that are previously altered. Then, as a second step, the developed model is applied to the actual data and the corresponding final results are reported. Analysis blinding allows researchers to plausibly show that the final results have not been p-hacked; similar to a pre-analysis plan. Data might be altered by not revealing treatment and control group or by permuting the outcome variable. Another solution could be data holdout models, which are built on one part of the data and later deployed on the other part of the data for inference. Note that this potentially trades the problem of data snooping for a statistical power problem (Moody et al. 2022). Still, in times of big data, the latter is likely to be much less of a problem and might therefore be preferred.

A very promising development is that we find increasing sample sizes and decreasing effect sizes, while p-values do not tend to become smaller. One interpretation is that larger sample sizes result in less p-hacking and less exaggerated effect sizes. Half of the previously published effect sizes have been shown to be exaggerated by a factor of two or more, and the true underlying effects might actually be much smaller or even economically irrelevant (Ioannidis et al. 2017).

Our analysis shows many positive developments in economics research that increase the reliability of published research and identifies areas for further improvement. It is crucial to recognize the credibility revolution and the open science movement for their positive contributions (Schiavone and Vazire 2023). While the replication crisis is discussed across disciplines,[18] the narrative of a crisis might be counterproductive and might lead to an indifferent generation of researchers. Rather, it should be conceived as a challenge (Fanelli 2018). A *Nature*'s survey with 1,576 researchers revealed that around 3 % neglect the existence of a crisis and 38 % state that there is only a slight crisis (Baker 2016). Consequently, actions to combat this crisis are not tilting at windmills. There is a realistic possibility that this crisis might be overcome with sufficient effort and training.

Limitations

This study is not without limitations. First, our sample has a structural break in 2016. The number of accessible journals drops substantially to 19 journals due to the dispute between the publisher *Elsevier* and the German universities.[19] Where possible, we analyzed the trends before and after the structural break, revealing at most only small differences. Second, we only analyze articles that are available in HTML. It cannot be ruled out that journals that only offer their articles in PDF format are written differently. Given that there is no inherent reason for these cases, we consider them unlikely. Third, our primary text-mining method is state-of-the-art in large-scale analysis (e.g., Currie et al. 2020) but limited in the sense that mentioning a method in the text does not necessarily imply that a method is actually used in the paper. However, trends in mentioning methods are

[18] Economists label the shift toward causal methods and improved econometrics as 'credibility revolution'. In contrast, psychologists label the replication crisis, the open science movement, and the credibility revolution as one movement, naming it the 'credibility revolution'. This perspective is driven by the notion that labeling these movements as one crisis might be misleading, given the numerous potential remedies for the issues in replication. Whereas our focus on the credibility revolution centers on methodological scrutiny as per Angrist and Pischke (2010), psychologists emphasize aspects such as preregistration and transparency, encompassing open science (Vazire 2018), which we consider a separate movement.

[19] For more information, consult, e.g., https://www.science.org/news/2016/12/thousands-german-researchers-set-lose-access-elsevier-journals, last retrieved on 31.08.2024.

likely to reflect corresponding usage of these methods. Fourth, the sample of empirical articles with statistical information excludes journals that permit the use of eye-catchers, potentially introducing a bias in the sample *empirical articles with statistical information*, as these journals are likely to be more aware of questionable research practices and transparency in reporting. Still, journals that permit the use of eye-catchers are an exception.

Future research
Future research could extend this study journal-wise or time-wise. This might be especially fruitful as recent artificial intelligence tools such as ChatGPT might have a measurable effect on scientific writing and, hence, on the results of text mining techniques. In addition, a third wave of empiricism is to be expected due to big data and machine learning (Lybbert and Buccola 2021), and the more recent advances in large language models. It will be important to take stock of the contributions made by the open science movement in a few years. As the critique on the first empirical wave and the replication crisis would not have been noted without the efforts of the growing field of meta-research (Vazire and Nosek 2023), it deserves the full attention and support of the whole research community.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505. DOI: 10.1198/jasa.2009.ap08746.

Abadie, Alberto and Javier Gardeazabal (2003). "The economic costs of conflict: A case study of the Basque Country". In: *American economic review* 93.1, pp. 113–132.

Andrews, Isaiah and Maximilian Kasy (2019). "Identification of and Correction for Publication Bias". In: *American Economic Review* 109.8, pp. 2766–2794. DOI: 10.1257/aer.20180310.

Angrist, Joshua, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu (2017). "Economic Research Evolves: Fields and Styles". In: *American Economic Review* 107.5, pp. 293–297. DOI: 10.1257/aer.p20171117.

Angrist, Joshua D. and Victor Lavy (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement*". In: *The Quarterly Journal of Economics* 114.2, pp. 533–575. DOI: 10.1162/003355399556061.

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". In: *Journal of Economic Perspectives* 24.2, pp. 3–30. DOI: 10.1257/jep.24.2.3.

Ankel-Peters, Jörg, Nathan Fiala, and Florian Neubauer (2025). "Is economics self-correcting? Replications in the American Economic Review". In: *Economic Inquiry* 63.2, pp. 463–485.

Askarov, Zohid, Anthony Doucouliagos, Chris Doucouliagos, and Tom D. Stanley (2023). "The Significance of Data-Sharing Policy". In: *Journal of the European Economic Association* 3.21, pp. 1191–1226. DOI: 10.1093/jeea/jvac053.

Bach, Laurent, Antoine Bozio, Arthur Guillouzouic, and Clément Malgouyres (2023). "Dividend Taxes and the Allocation of Capital: Comment". In: *American Economic Review* 113.7, pp. 2048–2052. DOI: 10.1257/aer.20221432.

Baker, Monya (2016). "1,500 scientists lift the lid on reproducibility". In: *Nature* 533.7604, pp. 452–454. DOI: 10.1038/533452a.

Bartoš, František, Maximilian Maier, Eric-Jan Wagenmakers, Franziska Nippold, Hristos Doucouliagos, et al. (2024). "Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics". In: *Research Synthesis Methods* 15.3, pp. 500–511.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1, pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-In-Differences Estimates?" In: *The Quarterly Journal of Economics* 119.1, pp. 249–275. DOI: 10.1162/003355304772839588.

Blanco-Perez, Cristina and Abel Brodeur (2020). "Publication Bias and Editorial Statement on Negative Findings". In: *The Economic Journal* 130.629, pp. 1226–1247. DOI: 10.1093/ej/ueaa011.

Briggs, Andrew H. (1999). "A Bayesian approach to stochastic cost-effectiveness analysis". In: *Health Economics* 8.3, pp. 257–261. DOI: 10.1002/(SICI)1099-1050(199905)8:3<257::AID-HEC427>3.0.CO;2-E.

Brodeur, Abel, Scott Carrell, David Figlio, and Lester Lusher (2023). "Unpacking p-Hacking and Publication Bias". In: *American Economic Review* 113.11, pp. 2974–3002. DOI: 10.1257/aer.20210795.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2020). "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics". In: *American Economic Review* 110.11, pp. 3634–3660. DOI: 10.1257/aer.20190687.

Brodeur, Abel, Nikolai Cook, and Carina Neisser (2024a). "P-hacking, data type and data-sharing policy". In: *The Economic Journal* 134.659, pp. 985–1018.

Brodeur, Abel, Nikolai M Cook, Jonathan S Hartley, and Anthony Heyes (2024b). "Do preregistration and preanalysis plans reduce p-hacking and publication bias? evidence from 15,992 test statistics and suggestions for improvement". In: *Journal of Political Economy Microeconomics* 2.3, pp. 527–561.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg (2016). "Star Wars: The Empirics Strike Back". In: *American Economic Journal: Applied Economics* 8.1, pp. 1–32. DOI: 10.1257/app.20150044.

Brodeur, Abel, Derek Mikola, Nikolai Cook, Lenka Fiala, Thomas Brailey, et al. (2026). "Reproducibility and robustness of economics and political science research". In: *Nature* 652.8108, pp. 151–156.

Bruns, Stephan B, Anthony Doucouliagos, Hristos Doucouliagos, Johannes König, Tom D Stanley, et al. (2025). "The delayed acceptance of female research in economics". In: *European Economic Review*, p. 105204.

Bruns, Stephan B, Christian Gross, and David I Stern (2014). "Is there really Granger causality between energy use and output?" In: *The Energy Journal* 35.4, pp. 101–134.

Bruns, Stephan B., Igor Asanov, Rasmus Bode, Melanie Dunger, Christoph Funk, et al. (2019). "Reporting errors and biases in published empirical findings: Evidence from innovation research". In: *Research Policy* 48.9. DOI: 10.1016/j.respol.2019.05.005.

Bruns, Stephan B., Teshome K. Deressa, Tom D. Stanley, Chris Doucouliagos, and John P. A. Ioannidis (2024a). "Estimating the extent of selective reporting: An application

to economics". In: *Research Synthesis Methods* 15.4, pp. 590–602. DOI: 10.1002/jrsm.1711.

Bruns, Stephan B., Helmut Herwartz, John P. A. Ioannidis, Chris-Gabriel Islam, and Fabian H. C. Raters (2023). "Statistical reporting errors in economics". In: *MetaArXiv*. DOI: 10.31222/osf.io/mbx62.

Bruns, Stephan B., Helmut Herwartz, Chris-Gabriel Islam, Thomas Kneib, and Robert Malina (2024b). "Ambiguous empirical results are not oversold but slightly less readable and more focused on statistical significance". In: mimeo.

Bruns, Stephan B. and John P. A. Ioannidis (2016). "p-Curve and p-Hacking in Observational Research". In: *PLOS ONE* 11. DOI: 10.1371/journal.pone.0149144.

Cala, P, T Havranek, Z Irsova, M Luskova, J Matousek, et al. (2026). "Financial incentives and performance: a meta-analysis of experiments in economics". In: *J Polit Econ Microecon forthcomming*.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, et al. (2016). "Evaluating replicability of laboratory experiments in economics". In: *Science* 351 (6280), pp. 1433–1436. DOI: 10.1126/science.aaf0918.

Card, David, Raj Chetty, Martin S. Feldstein, and Emmanuel Saez (2011). "Expanding Access to Administrative Data for Research in the United States". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.1888586.

Card, David and Stefano DellaVigna (2013). "Nine Facts about Top Journals in Economics". In: *Journal of Economic Literature* 51.1, pp. 144–161. DOI: 10.1257/jel.51.1.144.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri (2020). "Are Referees and Editors in Economics Gender Neutral?*". In: *The Quarterly Journal of Economics* 135.1, pp. 269–327. DOI: 10.1093/qje/qjz035.

Cardoso, Ana Rute, Paulo Guimarães, and Klaus F. Zimmermann (2010). "Trends in Economic Research: An International Perspective". In: *Kyklos* 63.4, pp. 479–494. DOI: 10.1111/j.1467-6435.2010.00484.x.

Cherrier, Beatrice (2017). "Classifying Economics: A History of the JEL Codes". In: *Journal of Economic Literature* 55.2, pp. 545–579. DOI: 10.1257/jel.20151296.

Chopra, Felix, Ingar Haaland, Christopher Roth, and Andreas Stegmann (2023). "The Null Result Penalty". In: *The Economic Journal* 134.657, pp. 193–219. DOI: 10.1093/ej/uead060.

Christensen, Garret, Allan Dafoe, Edward Miguel, Don A. Moore, and Andrew K. Rose (2019). "A study of the impact of data sharing on article citations using journal policies as a natural experiment". In: *PLOS ONE* 14.12. DOI: 10.1371/journal.pone.0225883.

Christensen, Garret and Edward Miguel (2018). "Transparency, Reproducibility, and the Credibility of Economics Research". In: *Journal of Economic Literature* 56.3, pp. 920–980. DOI: 10.1257/jel.20171350.

Collaboration, Open Science (2015). "Estimating the reproducibility of psychological science". In: *Science* 349.6251. DOI: 10.1126/science.aac4716.

Currie, Janet, Henrik Kleven, and Esmée Zwiers (2020). "Technology and Big Data Are Changing Economics: Mining Text to Track Methods". In: *AEA Papers and Proceedings* 110, pp. 42–48. DOI: 10.1257/pandp.20201058.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson (1986). "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project". In: *American Economic Review* 76, pp. 587–603.

Doucouliagos, Chris (2011). "How Large is Large? Preliminary and relative guidelines for interpreting partial correlations ineconomics". In: *School Working Paper, Faculty of Business and Law, Deakin University Australia*.

Doucouliagos, Hristos and Martin Paldam (2008). "Aid effectiveness on growth: A meta study". In: *European journal of political economy* 24.1, pp. 1–24.

Duflo, Esther, Abhijit Banerjee, Amy Finkelstein, Lawrence F. Katz, Benjamin A. Olken, et al. (2020). "In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plan for RCTs in Economics". In: *NBER Working Papers*.

Dunn, Olive Jean (1961). "Multiple Comparisons among Means". In: *Journal of the American Statistical Association* 56.293, pp. 52–64. DOI: 10.1080/01621459.1961.10482090.

Einav, Liran and Jonathan Levin (2014). "The Data Revolution and Economic Analysis". In: *Innovation Policy and the Economy* 14, pp. 1–24. DOI: 10.1086/674019.

Elliott, Graham, Nikolay Kudrin, and Kaspar Wüthrich (2022). "Detecting p-Hacking". In: *Econometrica* 90.2, pp. 887–906. DOI: 10.3982/ecta18583.

Fanelli, Daniele (2018). "Is science really facing a reproducibility crisis, and do we need it to?" In: *Proceedings of the National Academy of Sciences* 115.11, pp. 2628–2631. DOI: 10.1073/pnas.1708272114.

Ferguson, Joel, Rebecca Littman, Garret Christensen, Elizabeth L. Paluck, Nicholas Swanson, et al. (2023). "Survey of open science practices and attitudes in the social sciences". In: *Nature Communications* 14.1. DOI: 10.1038/s41467-023-41111-1.

Fire, Michael and Carlos Guestrin (2019). "Over-optimization of academic publishing metrics: observing Goodhart's Law in action". In: *GigaScience* 8.6. DOI: 10.1093/gigascience/giz053.

Garg, Prashant and Thiemo Fetzer (2025). "Causal Claims in Economics". In: *arXiv*. DOI: 10.48550/arxiv.2501.06873.

Goldsmith-Pinkham, Paul (2024). "Tracking the Credibility Revolution across Fields". In: *arXiv*. DOI: 10.48550/arxiv.2405.20604.

Gopalan, Maithreyi, Kelly Rosinger, and Jee Bin Ahn (2020). "Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges". In: *Review of Research in Education* 44.1, pp. 218–243. DOI: 10.3102/0091732x20903302.

GROBID (2023). *GROBID*. URL: https://github.com/kermitt2/grobid, last retrieved on 31.08.2024.

Hamermesh, Daniel (2025). "The Demographic and Research Styles of Economics Writing". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.5241501.

Hamermesh, Daniel S. (2013). "Six Decades of Top Economics Publishing: Who and How?" In: *Journal of Economic Literature* 51.1, pp. 162–172. DOI: 10.1257/jel.51.1.162.

Hardwicke, Tom E, Robert T Thibault, Jessica E Kosie, Joshua D Wallach, Mallory C Kidwell, et al. (2022). "Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017)". In: *Perspectives on Psychological Science* 17.1, pp. 239–251.

Hardwicke, Tom E, Joshua D Wallach, Mallory C Kidwell, Theiss Bendixen, Sophia Crüwell, et al. (2020). "An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017)". In: *Royal Society open science* 7.2.

Heckman, James J and Sidharth Moktan (2020). "Publishing and Promotion in Economics: The Tyranny of the Top Five". In: *Journal of Economic Literature* 58.2, pp. 419–470. DOI: 10.1257/jel.20191574.

Hengel, Erin (2022). "Publishing While Female: are Women Held to Higher Standards? Evidence from Peer Review". In: *The Economic Journal* 132.648, pp. 2951–2991. DOI: 10.1093/ej/ueac032.

Imai, Taisuke, Tom A Rutter, and Colin F Camerer (2021). "Meta-analysis of present-bias estimation using convex time budgets". In: *The Economic Journal* 131.636, pp. 1788–1814.

Imbens, Guido W. (2021). "Statistical Significance, p-Values, and the Reporting of Uncertainty". In: *Journal of Economic Perspectives* 35.3, pp. 157–174. DOI: 10.1257/jep.35.3.157.

Imbens, Guido W., Ran Spiegler, and Chris Taber (2019). *Statement from the editors of Econometrica, Quantitative Economics, and Theoretical Economics*. URL: https://www.econometricsociety.org/publications/statement-editors-econometrica-quantitative-economics-and-theoretical-economics, last retrieved on 31.08.2024.

Ioannidis, John P. A., Tom D. Stanley, and Chris Doucouliagos (2017). "The Power of Bias in Economics Research". In: *The Economic Journal* 127, pp. 236–265. DOI: 10.1111/ecoj.12461.

Islam, Chris-Gabriel (2024). "Meta-Research in times of Data Science. Using Automation, Machine Learning and modern Econometrics to analyze Open Science in Economics and related fields". PhD thesis. University of Göttingen and Hasselt University.

Kennedy, Peter E. (2002). "Sinning in the Basement: What are the Rules? The Ten Commandments of Applied Econometrics". In: *Journal of Economic Surveys* 16.4, pp. 569–589. DOI: 10.1111/1467-6419.00179.

Kerr, Norbert L. (1998). "HARKing: Hypothesizing After the Results are Known". In: *Personality and Social Psychology Review* 2.3, pp. 196–217. DOI: 10.1207/s15327957pspr0203_4.

Kocher, Martin G. and Matthias Sutter (2001). "The Institutional Concentration of Authors in Top Journals of Economics During the Last Two Decades". In: *The Economic Journal* 111.472, pp. 405–421. DOI: 10.1111/1468-0297.00637.

Leamer, Edward E. (1983). "Let's Take the Con Out of Econometrics". In: *American Economic Review*.

Lutmar, Carmela and Yaniv Reingewertz (2021). "Academic in-group bias in the top five economics journals". In: *Scientometrics* 126.12, pp. 9543–9556. DOI: 10.1007/s11192-021-04174-9.

Lybbert, Travis J. and Steven T. Buccola (2021). "The evolving ethics of analysis, publication, and transparency in applied economics". In: *Applied Economic Perspectives and Policy* 43.4, pp. 1330–1351. DOI: 10.1002/aepp.13131.

Matray, Adrien and Charles Boissel (2023). "Retraction of 'Dividend Taxes and the Allocation of Capital'". In: *American Economic Review* 113.7, pp. 2053–2054. DOI: 10.1257/aer.113.7.2053.

McCloskey, Deirdre N. and Stephen T. Ziliak (1996). "The Standard Error of Regressions". In: *Journal of Economic Literature* XXXIV, pp. 97–114. URL: https://www.jstor.org/stable/2729411, last retrieved on 31.08.2024.

Miguel, Edward (2021). "Evidence on research transparency in economics". In: *Journal of Economic Perspectives* 35.3, pp. 193–214.

Miske, Olivia, Anna Lou Abatayo, Mason Daley, Mirka Dirzo, Nicholas Fox, et al. (2026). "Investigating the reproducibility of the social and behavioural sciences". In: *Nature* 652.8108, pp. 126–134.

Moody, James W., Lisa A. Keister, and Maria C. Ramos (2022). "Reproducibility in the Social Sciences". In: *Annual Review of Sociology* 48.1, pp. 65–85. DOI: 10.1146/annurev-soc-090221-035954.

Mueller-Langer, Frank, Benedikt Fecher, Dietmar Harhoff, and Gert G. Wagner (2019). "Replication studies in economics—How many and which papers are chosen for replication, and why?" In: *Research Policy* 48.1, pp. 62–83. DOI: 10.1016/j.respol.2018.07.019.

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, et al. (2017). "A manifesto for reproducible science". In: *Nature Human Behaviour* 1.1. DOI: 10.1038/s41562-016-0021.

Munafò, Marcus R. and George D. Smith (2018). "Robust research needs many lines of evidence". In: *Nature* 553.7689. DOI: 10.1038/d41586-018-01023-3.

Nuijten, Michèle B., Chris H. J. Hartgerink, Marcel A. L. M. van Assen, Sacha Epskamp, and Jelte M. Wicherts (2016). "The prevalence of statistical reporting errors in psychology (1985–2013)". In: *Behavior Research Methods* 48, pp. 1205–1226. DOI: 10.3758/s13428-015-0664-2.

Olken, Benjamin A. (2015). "Promises and Perils of Pre-analysis Plans". In: *Journal of Economic Perspectives* 29.3, pp. 61–80. DOI: 10.1257/jep.29.3.61.

Poirier, Dale J. (2006). "The growth of Bayesian methods in statistics and economics since 1970". In: *Bayesian Analysis* 1.4, pp. 969–979. DOI: 10.1214/06-ba132.

Pütz, Peter and Stephan B. Bruns (2021). "The (non-)significance of reporting errors in economics: Evidence from three top journals". In: *Journal of Economic Surveys* 35.1, pp. 348–373. DOI: 10.1111/joes.12397.

Reid, Leonard N., Lawrence C. Soley, and Roger D. Winner (1981). "Replication in Advertising Research: 1977, 1978, 1979". In: *Journal of Advertising* 10.1, pp. 3–13. DOI: 10.1080/00913367.1981.10672750.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf (2010). "Hypothesis Testing in Econometrics". In: *Annual Review of Economics* 2.1, pp. 75–104. DOI: 10.1146/annurev.economics.102308.124342.

Rosenthal, Robert (1979). "The file drawer problem and tolerance for null results." In: *Psychological Bulletin* 86.3, p. 638. DOI: 10.1037/0033-2909.86.3.638.

Rubin, Mark and Chris Donkin (2022). "Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests". In: *Philosophical Psychology*, pp. 1–29. DOI: 10.1080/09515089.2022.2113771.

Rupp, Nicholas G. and Carl Nicholas McKinney Jr. (2002). "Publication Patterns of the Elite Economics Departments: 1995-2000". In: *Eastern Economic Journal* 28.4, pp. 523–538. URL: https://www.jstor.org/stable/40325396.

Sarafoglou, Alexandra, Suzanne Hoogeveen, and Eric-Jan Wagenmakers (2023). "Comparing Analysis Blinding With Preregistration in the Many-Analysts Religion Project". In: *Advances in Methods and Practices in Psychological Science* 6.1. DOI: 10.1177/25152459221128319.

Sayre, Franklin and Amy Riegelman (2018). "The Reproducibility Crisis and Academic Libraries". In: *College & Research Libraries* 79.1. DOI: 10.5860/crl.79.1.2.

Schiavone, Sarah R. and Simine Vazire (2023). "Reckoning With Our Crisis: An Agenda for the Field of Social and Personality Psychology". In: *Perspectives on Psychological Science* 18.3, pp. 710–722. DOI: 10.1177/17456916221101060.

Scott, Loren C. and Peter M. Mitias (1996). "Trend in ranking of economics departments in the U.S.: An update". In: *Economic Inquiry* 34.2, pp. 378–400. DOI: 10.1111/j.1465-7295.1996.tb01383.x.

Stanley, Tom D (2001). "Wheat from chaff: Meta-analysis as quantitative literature review". In: *Journal of economic perspectives* 15.3, pp. 131–150.

— (2005). "Beyond Publication Bias". In: *Journal of Economic Surveys* 19.3, pp. 309–345. DOI: 10.1111/j.0950-0804.2005.00250.x.

Stock, James H. (2010). "The Other Transformation in Econometric Practice: Robust Tools for Inference". In: *Journal of Economic Perspectives* 24.2, pp. 83–94. DOI: 10.1257/jep.24.2.83.

Swanson, Nicholas, Garret Christensen, Rebecca Littman, David Birke, Edward Miguel, et al. (2020). "Research transparency is on the rise in economics". In: *AEA Papers and Proceedings* 110, pp. 61–65.

Tyner, Andrew H, Anna Lou Abatayo, Mason Daley, Samuel Field, Nicholas Fox, et al. (2026). "Investigating the replicability of the social and behavioural sciences". In: *Nature* 652.8108, pp. 143–150.

UNESCO (2021). *UNESCO Recommendationon Open Science*. URL: https://unesdoc.unesco.org/ark:/48223/pf0000379949/PDF/379949eng.pdf.multi, last retrieved on 31.08.2024.

Vazire, Simine (2018). "Implications of the Credibility Revolution for Productivity, Creativity, and Progress". In: *Perspectives on Psychological Science* 13.4, pp. 411–417. DOI: 10.1177/1745691617751884.

Vazire, Simine and Brian Nosek (2023). "Introduction to Special Topic "Is Psychology Self-Correcting? Reflections on the Credibility Revolution in Social and Personality Psychology"". In: *Social Psychological Bulletin* 18, pp. 1–4. DOI: 10.32872/spb.12927.

Vivalt, Eva (2019). "Specification Searching and Significance Inflation Across Time, Methods and Disciplines". In: *Oxford Bulletin of Economics and Statistics* 81.4, pp. 797–816. DOI: 10.1111/obes.12289.

Vlaeminck, Sven (2021). "Dawning of a new age? Economics journals' data policies on the test bench". In: *LIBER Quarterly: The Journal of the Association of European Research Libraries* 31.1, pp. 1–29. DOI: 10.53377/lq.10940.

Wilkinson, Mark D., Michel Dumontier, IJsbrand J. Aalbersberg, Gabrielle Appleton, Myles Axton, et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3.1. DOI: 10.1038/sdata.2016.18.

Winkler, Anne E, Sharon G Levin, Paula E Stephan, and Wolfgang Glänzel (2014). "Publishing Trends in Economics across Colleges and Universities, 1991–2007". In: *Eastern Economic Journal* 40.4, pp. 560–582. DOI: 10.1057/eej.2013.29.

Ziliak, Stephen T. and Deirdre N. McCloskey (2004). "Size matters: the standard error of regressions in the American Economic Review". In: *The Journal of Socio-Economics* 33.5, pp. 527–546. DOI: 10.1016/j.socec.2004.09.024.

Zong, Qianjin, Zhihong Huang, and Jiaru Huang (2023). "Do open science badges work? Estimating the effects of open science badges on an article's social media attention and research impacts". In: *Scientometrics* 128.6, pp. 3627–3648. DOI: 10.1007/s11192-023-04720-7.

How did empirical economic research improve over the last decades: Recent trends and prospects

Online Appendix

Abel Brodeur[1,2,3], Stephan B. Bruns[4,5,6,†], Helmut Herwartz[7] and Chris-Gabriel Islam[4,7,8]

[1]University of Ottawa, Canada
[2]Institute for Replication
[3]IZA – Institute of Labor Economics, Germany
[4]Hasselt University, Belgium
[5]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, USA
[6]International Center for Higher Education Research (INCHER) Kassel, Germany
[7]Georg August University of Göttingen, Germany
[†]Corresponding author: stephan.bruns@uhasselt.be

August 17, 2026

Table of Contents

A Additional explanations **3**
A.1 Methods 3

B Tables **4**

C Figures **13**

List of Tables

B.1 Overview *all articles* 4
B.2 Overview *empirical articles* 5
B.3 Overview *empirical articles with statistical information* 6
B.4 Overview *articles with data and/or code* 7
B.5 Overview *matched articles with data and/or code* 8

List of Figures

C.1 Share of empirical articles mentioning "pre-analysis plan" (empirical articles: 22k) 13
C.2 Share of empirical articles mentioning "bootstrapping" (empirical articles: 22k) 13

A Additional explanations

A.1 Methods

The handling of footnotes in the full text of HTML documents is noisy, as some journals tend to have the footnotes in line next to the text, whereas other journals have their footnotes at the end of the HTML document. As most journals have footnotes located at the end of the HTML document, we were able to separate the full text from the footnotes and focus only on the full text. For an even more robust extraction and to avoid including figure or table captions, we exclude sentences without a verb. For the detection of verbs, we rely on the Natural Language Processing toolkit *SpaCy*. Additionally, we remove sentences that are clearly noise, e.g., *"Forgot your password?"* or *"Create a new account"*.

B Tables

Table B.1: Overview *all articles*

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Sum
ARE	0	0	0	0	0	0	0	0	0	0	0	21	25	25	23	19	36	24	22	22	22	35	274
ARFE	0	0	0	0	0	0	0	0	0	0	0	2	13	21	19	14	19	22	12	14	21	12	169
EER	68	88	102	98	90	52	68	91	91	89	63	61	58	52	102	102	118	110	153	0	0	0	1,656
EJ	0	0	0	0	73	76	76	80	82	98	97	87	81	77	70	73	79	91	83	120	131	106	1,580
EP	0	0	0	0	15	14	12	16	16	16	16	16	17	15	16	16	16	16	16	17	16	7	273
IER	0	0	0	0	44	54	48	59	47	54	48	45	48	51	55	50	50	47	55	44	78	67	944
JBES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	37	32	40	46	43	51	58	353
JEEA	0	0	0	0	0	0	0	0	0	0	0	0	0	41	51	51	54	36	43	38	47	51	412
JMCB	0	0	0	0	0	0	0	0	0	95	80	88	85	100	88	81	90	91	60	59	62	75	1,054
JME	40	49	52	51	62	64	63	62	96	124	93	85	71	43	54	57	72	77	67	0	0	0	1,282
JOAE	0	0	0	38	37	39	50	48	68	59	44	51	49	54	55	53	60	59	73	78	63	69	1,047
JODE	52	66	55	62	49	67	76	71	64	85	61	76	72	80	104	108	119	70	81	0	0	0	1,418
JOE	46	69	82	79	93	90	86	80	124	176	161	99	140	140	167	95	145	193	147	0	0	0	2,212
JOEEM	0	0	0	0	0	66	54	62	43	46	40	49	40	57	58	69	53	42	43	0	0	0	722
JOES	0	0	0	0	0	0	0	0	15	28	29	33	35	39	39	42	46	44	43	56	55	59	563
JOF	78	76	93	88	89	93	90	86	87	84	81	77	69	60	60	67	71	70	71	60	65	73	1,688
JOFE	45	55	56	61	58	60	75	79	87	103	96	93	100	136	124	153	102	118	123	0	0	0	1,724
JOFI	0	0	0	0	0	15	18	18	23	21	23	27	25	28	28	32	24	27	24	0	0	0	333
JOIE	46	54	48	62	59	58	59	70	62	66	79	72	66	62	87	98	79	89	101	0	0	0	1,317
JOIMF	49	38	44	47	43	46	50	63	64	65	77	70	96	94	105	143	121	114	116	0	0	0	1,445
JOLE	0	0	0	24	41	33	32	28	27	24	21	20	26	23	25	31	26	38	48	38	37	50	592
JOUE	0	0	0	0	0	0	0	0	0	0	0	20	55	34	46	46	53	41	48	0	0	0	343
JPE	0	0	26	44	48	42	56	42	38	30	31	30	30	30	30	31	30	32	39	44	67	46	766
JPUE	90	80	73	85	65	117	124	97	108	101	136	101	94	136	92	106	136	116	84	0	0	0	1,941
OBES	0	0	0	0	25	41	39	41	45	37	39	38	36	39	41	46	44	44	42	48	50	60	755
QJE	0	0	0	0	0	0	0	0	0	0	0	0	0	46	41	40	40	40	40	40	40	10	337
RAND	0	0	0	0	0	0	0	0	0	30	51	34	34	32	32	32	35	36	39	42	36	36	469
RED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	50	50	0	0	0	134
REEP	0	0	0	0	0	0	0	0	0	16	17	17	17	16	16	16	16	16	18	21	0	0	186
RESTUD	0	0	0	0	0	20	48	46	43	47	48	50	50	50	52	52	52	48	51	52	66	76	851
RFS	28	41	36	38	54	38	37	39	41	59	81	147	119	108	95	84	93	87	91	112	120	123	1,671
RP	83	46	65	84	81	101	93	94	98	100	126	133	110	114	132	133	134	140	150	0	0	0	2,017
WBRO	0	0	0	0	0	0	0	5	10	10	10	10	12	13	10	11	10	10	10	8	8	10	147
Sum	625	662	732	861	1,026	1,186	1,254	1,277	1,379	1,663	16,48	1,653	16,73	1,816	1,963	1,988	2,089	2,078	2,089	957	1,035	1,023	30,675

Table B.2: Overview *empirical articles*

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Sum
ARE	0	0	0	0	0	0	0	0	0	0	0	9	8	7	16	12	17	9	13	10	12	9	122
ARFE	0	0	0	0	0	0	0	0	0	0	0	0	7	13	7	4	4	9	7	7	10	9	77
EER	27	40	36	54	41	27	32	52	47	60	44	36	40	34	74	81	91	91	131	0	0	0	1,038
EJ	0	0	0	0	42	50	54	53	53	67	67	60	63	54	58	63	67	62	71	98	107	88	1,177
EP	0	0	0	0	14	12	12	15	15	16	16	16	16	15	16	16	16	16	15	16	16	7	265
IER	0	0	0	0	16	29	22	14	21	27	25	31	22	35	33	32	30	37	34	21	53	48	530
JBES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	37	32	40	44	41	46	56	340
JEEA	0	0	0	0	0	0	0	0	0	0	0	0	0	26	41	38	36	30	36	32	42	45	326
JMCB	0	0	0	0	0	0	0	0	0	71	62	66	67	63	64	63	66	65	50	50	56	66	809
JME	29	30	40	37	46	41	48	43	67	100	67	69	57	34	49	50	62	67	57	0	0	0	993
JOAE	0	0	0	33	30	34	49	46	59	56	39	48	43	48	54	53	56	56	67	76	59	64	970
JODE	37	47	40	40	38	51	54	52	50	57	49	62	63	68	95	94	103	62	68	0	0	0	1,130
JOE	41	57	71	62	75	74	64	68	110	137	139	84	116	123	136	85	130	172	130	0	0	0	1,874
JOEEM	0	0	0	0	0	41	32	29	23	29	25	29	25	41	36	50	42	27	33	0	0	0	462
JOES	0	0	0	0	0	0	0	0	1	14	16	23	21	14	19	22	27	29	28	30	46	42	332
JOF	67	71	73	65	82	77	76	74	76	76	68	67	60	50	52	63	61	62	64	55	58	62	1,459
JOFE	41	52	48	57	50	57	68	73	72	93	91	84	94	129	114	140	98	105	118	0	0	0	1,584
JOFI	0	0	0	0	0	11	11	13	16	15	13	20	19	19	24	22	17	24	19	0	0	0	243
JOIE	25	31	23	31	31	34	42	42	39	41	60	48	48	49	71	77	66	81	90	0	0	0	929
JOIMF	42	30	35	37	38	35	43	56	58	59	66	65	86	88	99	128	120	110	106	0	0	0	1,301
JOLE	0	0	0	18	33	29	30	23	21	22	19	19	24	22	21	29	25	35	45	36	34	46	531
JOUE	0	0	0	0	0	0	0	0	0	0	0	16	46	27	37	39	43	35	47	0	0	0	290
JPE	0	0	17	29	38	26	38	27	22	18	22	25	21	21	19	16	21	27	31	38	53	37	546
JPUE	35	42	39	43	42	52	78	65	60	66	83	72	70	99	61	82	101	89	68	0	0	0	1,247
OBES	0	0	0	0	25	37	34	39	38	36	37	37	35	38	40	44	40	43	40	46	50	54	713
QJE	0	0	0	0	0	0	0	0	0	0	0	0	0	36	32	29	34	33	35	38	36	10	283
RAND	0	0	0	0	0	0	0	0	0	11	15	8	14	12	11	15	18	17	18	17	15	14	185
RED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	40	37	0	0	0	104
REEP	0	0	0	0	0	0	0	0	0	8	5	5	8	12	7	6	6	7	12	9	0	0	85
RESTUD	0	0	0	0	0	0	0	0	0	0	0	0	0	25	33	34	38	27	37	31	50	56	331
RFS	17	25	20	29	33	30	25	30	36	45	72	131	105	89	84	71	77	74	85	104	117	106	1,405
WBRO	0	0	0	0	0	0	0	2	6	6	9	6	9	6	4	5	7	5	8	7	6	8	94
Sum	361	425	442	535	674	747	812	816	890	1,130	1,109	1,136	1,187	1,297	1,451	1,500	1,578	1,586	1,644	762	866	827	21,775

Table B.3: Overview *empirical articles with statistical information*

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Sum
ARE	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	1	4
ARFE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
EER	0	3	0	3	3	5	6	7	9	10	9	7	6	11	20	17	27	37	49	229
EJ	0	0	0	0	0	4	7	9	7	13	15	14	27	13	13	18	19	29	36	224
EP	0	0	0	0	4	2	2	3	2	5	9	7	6	5	9	3	6	6	7	76
IER	0	0	0	0	0	0	0	0	0	1	1	2	1	3	3	4	1	4	5	25
JEEA	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	1	16	12	11	48
JMCB	0	0	0	0	0	0	0	0	0	7	7	14	19	7	5	15	5	8	8	95
JME	0	0	1	1	1	3	5	3	5	3	2	2	5	2	1	2	8	7	5	56
JOAE	0	0	0	1	2	2	2	1	2	2	2	0	2	6	3	2	4	0	4	35
JODE	2	3	3	5	8	7	6	10	13	11	27	23	22	24	46	46	56	44	39	395
JOE	0	1	0	0	0	0	1	0	0	2	4	0	3	1	2	0	4	3	2	23
JOEEM	0	0	0	0	0	5	5	5	0	4	4	6	8	15	7	16	20	8	11	114
JOES	0	0	0	0	0	0	0	0	0	0	0	1	0	1	3	1	0	1	1	8
JOF	6	7	10	5	10	3	10	16	14	11	17	18	17	15	25	26	17	26	35	288
JOFE	4	1	5	5	7	9	10	8	15	25	22	23	27	37	31	39	39	43	46	396
JOFI	0	0	0	0	0	0	2	0	2	0	1	0	7	4	11	12	6	12	13	70
JOIE	3	2	0	2	2	3	8	8	10	7	14	9	14	20	24	27	24	38	36	251
JOIMF	1	1	1	0	1	1	3	5	4	7	6	13	13	7	26	31	25	25	36	206
JOLE	0	0	0	3	9	5	4	8	4	7	5	9	6	4	9	14	12	16	19	134
JOUE	0	0	0	0	0	0	0	0	0	0	0	4	16	11	16	15	22	17	17	118
JPE	0	0	3	3	4	3	8	4	6	5	5	6	9	7	6	5	7	6	11	98
JPUE	3	4	2	5	4	5	11	14	15	22	15	27	20	35	20	39	42	43	32	358
OBES	0	0	0	0	1	4	2	3	0	6	3	7	1	7	7	7	3	6	11	68
QJE	0	0	0	0	0	0	0	0	0	0	0	0	0	11	15	12	19	17	12	86
RAND	0	0	0	0	0	0	0	0	0	1	4	0	3	3	2	6	6	6	1	32
RED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	5
REEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
RESTUD	0	0	0	0	0	0	0	0	0	0	0	0	0	3	10	9	11	12	11	56
RFS	0	1	5	1	3	4	3	1	3	6	7	29	37	25	22	19	27	23	29	245
WBRO	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Sum	19	23	30	34	59	65	95	105	111	155	179	221	270	283	340	387	428	453	489	3,746

Table B.4: Overview *articles with data and/or code*

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023		
AEJAPP	0	0	0	0	35	40	37	40	38	40	35	37	42	46	46	43	37	58	38	612	
AEJMAC	0	0	0	0	20	38	32	31	29	31	35	30	32	32	37	33	41	50	39	510	
AEJMIC	0	0	0	0	23	33	38	32	34	44	49	43	43	40	23	13	29	34	16	494	
AEJPOL	0	0	0	0	19	31	30	37	43	45	46	39	52	48	52	51	54	63	46	656	
AER	196	202	208	208	226	247	270	188	248	272	258	274	273	114	105	93	99	93	60	3,634	
AERI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	25	24	24	18	108	
ECTA	0	0	0	0	0	0	0	0	0	0	16	27	35	34	25	44	57	55	24	317	
JAERE	0	0	0	0	0	0	0	0	0	0	0	0	0	14	17	35	31	48	26	171	
JEEA	0	0	0	0	0	0	0	0	0	11	22	31	22	29	36	57	29	50	13	300	
JEP	0	0	0	0	0	58	52	50	53	52	49	49	48	48	27	18	25	26	21	576	
JOLE	0	0	0	0	0	0	0	0	0	0	22	25	27	25	29	29	29	16	8	210	
JPE	0	0	0	0	0	0	0	0	0	0	33	39	34	46	47	39	36	28	6	308	
MS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	44	64	202	130	467	
PANDP	0	0	0	0	0	0	0	0	0	0	0	0	0	84	81	86	76	52	88	467	
QJE	0	0	0	0	0	0	0	0	0	0	0	0	35	30	37	41	46	48	22	259	
RESTAT	0	0	0	0	0	39	90	75	120	67	79	66	75	129	79	114	135	121	57	1,246	
RESTUD	0	0	0	0	0	52	50	52	53	52	48	52	56	64	74	70	70	76	36	805	
Sum	196	202	208	208	323	538	599	505	618	614	692	712	774	783	759	835	882	1,044	648	11,140	34	22	486

Table B.5: Overview *matched articles with data and/or code*

	2011	2012	2013	2014	2015	2016	2017	2018	2019	Sum
JEEA	26	41	38	36	30	36	32	42	45	326
JOLE	22	21	29	25	35	45	36	34	46	293
JPE	21	19	16	21	27	31	38	53	37	263
QJE	36	32	29	34	33	35	38	36	10	283
RESTUD	25	33	34	38	27	37	31	50	56	331
Sum	130	146	146	154	152	184	175	215	194	1,496

Topic	Regular Expression	Case Sensitive	Origin
administrative data	r'(administrative data|admin data|administrative-data|admin-data|administrative record|admin record|administrative regist|admin regist|register data|registry data)'	N	C
Bayes	r'bayes'	N	O
bootstrapping	bootstrap	N	O
bunching	r'bunching'	N	C
clustered inference	r'clustered|clustering|clusters?\s(over|at|by)|cluster\sstandard|allowing\s(for\s)?correlation\sat(?:\sthe)?\s[\w-]*?\slevel|standard\serrors(?:\sin\sparentheses,)?\scorrected\s(for|by)(?!\sheterosk)(?!\sautocorre)|standard\serrors(?:\sin\sparentheses,)?\sadjusted\sfor\s([\w-]*?\s)?cluster|cluster.robust|cluster\serrors\sby|intra\-cluster\scorrelation|cluster.correct|obs\w+\sper\scluster|cluster\s(\([^()]*?\))?\sand\sheteroscedasticity\protect \discretionary {\char \hyphenchar \font }{}{}corrected\sstandard\serrors'		
confidence intervals	r'\b(confidence.interval)'	N	C*
data	r'data'	N	C
difference-in-differences	r'Difference in Diff|Difference in diff|difference in diff|Difference-in-Diff|Difference-in-diff|difference-in-diff|Differences in Diff|Differences in diff|differences in diff|Differences-in-Diff|Differences-in-diff|differences-in-diff|diff-in-diff|d-in-d|DiD'	Y	C
effect size	r'(effect|coefficient)(.s)?\s(size|magnitude|level)|(size|magnitude|level)\sof\s(effect|coefficient)'	N	O
elasticity	r'\b(elasticit)'	N	O
event study	r'event stud|event-stud'	N	C
exploratory	r'exploratory|explorative|hypothes[ie]s[\s-]generating|hypothes[ie]s[\s-]generation'	N	O
field experiment	r'field.experiment'	N	C
fixed effects	r'FE|Fixed Effect|Fixed effect|fixed effect|Fixed Effects|Fixed effects|fixed effects|Fixed-Effect|Fixed-effect|fixed-effect|Fixed-Effects|Fixed-effects|fixed-effects'	Y	C
hypothesis testing	r'hypothes[ie]s[\s-]test(ing|ed)?|test(ing|ed|s?\sof)?\s((a|the)\s)?hypothes[ie]s|hypothes[ie]s\s((that|which)\s)?(is|are)\stested'	N	O

Continued on next page

Table B.6 – *Continued from previous page*

Topic	Regular Expression	Case Sensitive	Origin
identification	r'identif\w+(\s\w+){0,2}\s(effect|response|impact|elastic it|parameter|coeficient)(?!ive);causal identification|causal ly identified|identification strategy|identification approach| identification assumption|identifying assumption|identifyin g variation|empirical identification|over identified|over-ide ntified|under identified|under-identified|identification pro perties|identification test|identification problem|identificat ion issue|problem with identification|problems with identific ation|issue with identification|issues with identification|pr oblem identifying|problems identifying|issue identifying ;is sues identifying|threat to identification|threats to identificat ion|threat for identification|threats for identification|over id entifying|over-identifying|under identifying|under-identifyi ng|partial identification|partially identified|non-parametri c identification|nonparametric identification|non parametri c identification|non-parametrically identified|nonparamet rically identified|non parametrically identified|identificatio n condition|identifying condition|condition for identificatio n|conditions for identification|condition for identifying|cond itions for identifying|point identification|point-identificatio n|point identified|point-identified|point identifying|point-i dentifying set identification|set-identification|set identified| set-identified|set identifying |set-identifying|identification a nalysis|weak identification|identification result|identificatio n argument|identification framework|identification scheme'	N	C*
instrumental variable	r'Instrumental Variable|Instrumental variable|instrument al variable|Instrumental-Variable|Instrumental-variable|in strumental-variable|Two Stage Least Squares|Two stage lea st squares|two stage least squares|2SLS|TSLS|valid instru ment|exogenous instrument|IV Estimat|IV estimat|IV-estim at|IV Specification|IV specification|IV-specification|IV Re gression|IV regression|IV-regression|IV Strateg|IV strateg| IV-strateg|we instrument|I instrument|paper instruments|e xclusion restriction|weak first stage|simulated instrument'	Y	C
internet data	r'internet data|internet-data|web data|web-data|scraped d ata|scraped-data|scrape data|scraping data|search data|s earch-data|google data|google-data|social media data|goog le trend|google-trend|google search|google-search|google ng ram|google n-gram|google books ngram|google books n-gra m'	N	C
interview data	r'interview.data'	N	C
lab experiment	r'Laboratory Experiment|Laboratory experiment|laborato ry experiment|Lab Experiment|Lab experiment|lab experi ment|Dictator Game|Dictator game|dictator game|Ultima tum Game|Ultimatum game|ultimatum game|Trust Game| Trust game|trust game|Public Good Game|Public good g ame|public good game|Public Goods Game|Public goods ga me|public goods game|Z-tree|zTree|ORSEE|show-up fee|l aboratory participant|lab participant'	Y	C

Continued on next page

Table B.6 – *Continued from previous page*

Topic	Regular Expression	Case Sensitive	Origin
matching	r'propensity score|propensity score matching|propensity-score matching|matching estimat|nearest neighbor matching|nearest-neighbor matching|nearest neighbour matching|nearest-neighbour matching|caliper matching|stratification matching|exact matching|one to one matching|one-to-one matching|kernel matching|inverse probability matching|inverse-probability matching'	N	C
multiple hypothesis testing	r'multiple\s(?:testing|comparison|p.value)|adjusted\saccording\sto\sbonferroni|bonferroni.(correct|adjusted)|\sfwer\s|(?:family|experiment)wise\serror\srate|holm.bonferroni.method|[Åăs]id[aÃą]k.correction|closed.testing.procedure|boole.bonferroni.bound|duncan(?:.s)?\snew\smultiple\srange\stest|harmonic\smean\sp.value\sprocedure|tukey(?:.s)?\srange\stest|hochberg(?:.s)?\sstep.up\sprocedure|dunnett(?:.s)?\stest|scheff[eÃľ](?:.s)?\smethod|\sfdr\s|false\sdiscovery\srate|huynh[\s-]feld'	N	O
null findings	r'null[\s-](result|finding|effect)'	N	O
p-hacking	r'p.hack|data.dredg|data.snoop|selective.report|data.fiddling|significance.chas'	N	O
proprietary data	r'proprietary data|confidential data|nonpublic data|non-public data|proprietary-data|confidential-data|nonpublic-data|non-public-data'	N	C
pre-analysis plan	r'pre-analysis plan|pre analysis plan|preanalysis plan'	N	C
publication bias	r'publication\sbias'	N	O
quasi- and natural experiments	r'quasi experiment|quasi-experiment|quasiexperiment|natural experiment|natural-experiment'	N	C
randomized controlled trial	r'Randomized Controlled Trial|Randomized controlled trial|randomized controlled trial|Randomized Control Trial|Randomized control trial|randomized control trial|Randomized Field Experiment|Randomized field experiment|randomized field experiment|Randomized Controlled Experiment|Randomized controlled experiment|randomized controlled experiment|Randomised Controlled Trial|Randomised controlled trial|randomised controlled trial|Randomised Control Trial|Randomised control trial|randomised control trial|Randomised Field Experiment|Randomised field experiment|randomised field experiment|Randomised Controlled Experiment|Randomised controlled experiment|randomised controlled experiment|Social Experiment|Social experiment|social experiment|RCT'	Y	C
regression discontinuity	r'Regression Discontinuit|Regression discontinuit|regression discontinuit|Regression-discontinuity|regression-discontinuity|Regression Kink|Regression kink|regression kink|RD Design|RD design|RD-design|RD Estimat|RD estimat|RD-estimat|RD Model|RD model|RD-model|RD Regression|RD regression|RD-regression|RD Coefficient|RD coefficient|RD-coefficient|RK Design|RK design|RK-Design|RK-design|RKD|RDD'	Y	C
replication	r'reproducibility|replication'	N	O
retraction	r'retraction|retracted'	N	O

Continued on next page

Table B.6 – *Continued from previous page*

Topic	Regular Expression	Case Sensitive	Origin
semi- and non-parametric methods	r'semi.?parametric|non.?parametric'	N	O
specification searching	r'specification\ssearch'	N	O
statistical power	r'statistical\spower|g[*^\s-]{1,2}power|power\sanalysis'	N	O
survey data	r'survey[^.!?]*?data;data[^.!?]survey'	N	C*
synthetic control method	r'synthetic control'	N	C

Notes: Letter C in the origin column means that the regular expression is directly applied from Currie et al. (2020) while C* means that the expression is based on the expression by Currie et al. (2020). Letter O means that the regular expression is written by ourselves.

C Figures

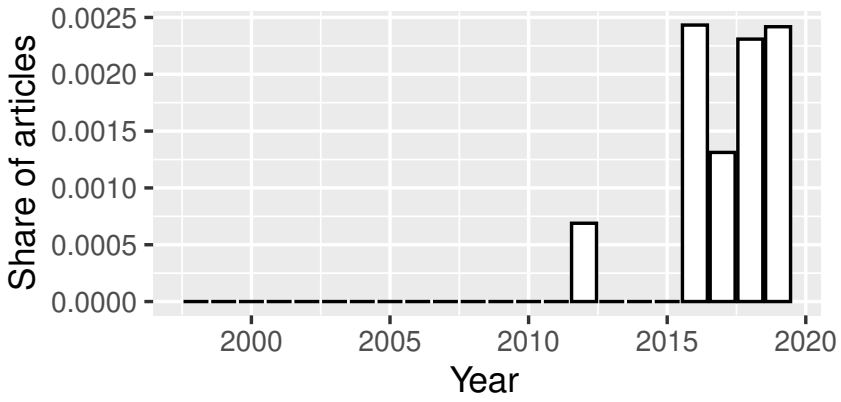


Figure C.1: Share of empirical articles mentioning "pre-analysis plan" (empirical articles: 22k)

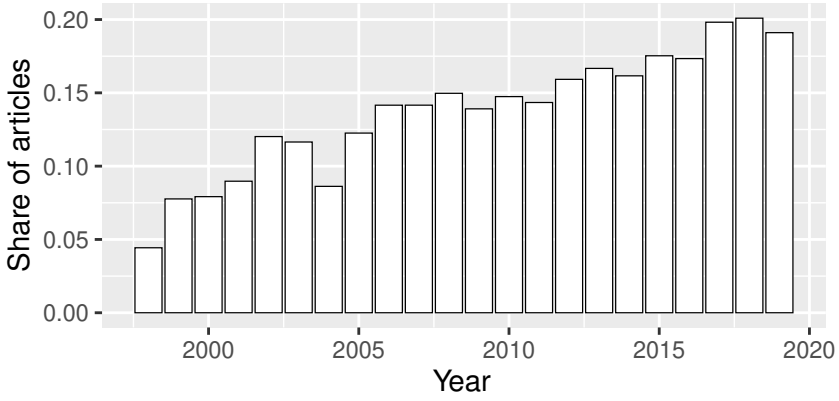


Figure C.2: Share of empirical articles mentioning "bootstrapping" (empirical articles: 22k)