

Ankel-Peters, Jörg; Fiala, Nathan; Neubauer, Florian

Working Paper

Is economics self-correcting? Replications in the American Economic Review

Ruhr Economic Papers, No. 1005

Provided in Cooperation with:

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Ankel-Peters, Jörg; Fiala, Nathan; Neubauer, Florian (2023) : Is economics self-correcting? Replications in the American Economic Review, Ruhr Economic Papers, No. 1005, ISBN 978-3-96973-171-0, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, <https://doi.org/10.4419/96973171>

This Version is available at:

<https://hdl.handle.net/10419/269235>

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.



RUHR

ECONOMIC PAPERS

Jörg Ankel-Peters
Nathan Fiala
Florian Neubauer

Is Economics Self-Correcting? Replications in the *American Economic Review*

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt,
Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #1005

Responsible Editor: Manuel Frondel

All rights reserved. Essen, Germany, 2023

ISSN 1864-4872 (online) – ISBN 978-3-96973-171-0

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #1005

Jörg Ankel-Peters, Nathan Fiala and Florian Neubauer

**Is Economics Self-Correcting?
Replications in the
*American Economic Review***

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

<http://dx.doi.org/10.4419/96973171>

ISSN 1864-4872 (online)

ISBN 978-3-96973-171-0

Jörg Ankel-Peters, Nathan Fiala and Florian Neubauer¹

Is Economics Self-Correcting? Replications in the *American Economic Review*

Abstract

Replication and constructive controversy are essential for scientific progress. This paper reviews the impact of all replications published as comments in the American Economic Review between 2010 and 2020. We investigate the citation rates of comments and whether a comment affects its original paper's citation rates. We find that most comments are barely cited, and they have no impact on the original papers' subsequent citations. This finding holds for original papers for which the comment diagnoses a substantive problem. We conclude from these citation patterns that replications do not update the economics literature. In an online opinion survey, we elicited viewpoints of both comment authors and original authors and find that in most cases, there is no consensus regarding the replication's success and to what extent the original paper's contribution sustains. This resonates with the conventional wisdom that robustness and replicability are hard to define in economics.

JEL-Code: A11, A14

Keywords: Replication; citations; meta-science

February 2023

¹ Jörg Ankel-Peters, RWI and University of Passau; Nathan Fiala, University of Connecticut and RWI; Florian Neubauer, University of Connecticut and RWI. – Funding: This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) [Grant No. 3473/1-1 within the DFG Priority Program META-REP (SPP 2317)]. We are grateful for valuable comments and suggestions by Krisztina Kis-Katos. We also thank research seminar participants at the University of Wageningen and the University of Siegen. We acknowledge the participation of all AER editors since 1985 in a short e-mail survey: Orley Ashenfelter, Ben Bernanke, Robert Moffitt, Pinelopi Goldberg, Esther Duflo, and Erzo Luttmer. We further acknowledge the participation of the following people in our survey of authors of original papers, comments, and replies in our AER sample: Armon Rezai, Art Fraas, Brock Smith, Christoph March, Frank Venmans, Gary Bolton, Jens Iversen, Jérôme Adda, Johannes Pfeijfer, John William Hatfield, Jonathan Ketcham, Jordi Galí, Juan Rubio-Ramírez, Julian Wright, Karel Mertens, Kathryn Zeiler, Kurt G. Lunsford, Marcel Timmer, Matteo Cervellati, Miklós Koren, Paul Milgrom, Rafael Di Tella, Samuel Lundstrom, Sebastian Claro, Sweta Chaman Saxena, Uwe Sunde, Yuval Heller, Zhi Wang, Zhouxiang Shen as well as 69 authors who participated in our survey but preferred not to be acknowledged. Caitlin Chan, Rita Reischl, Youmna Eid, and Martin Buchner provided valuable research assistance. – All correspondence to: Jörg Ankel-Peters, RWI – Leibniz Institute for Economic Research, Hohenzollernstr. 1-3, 45128 Essen, Germany; Email: peters@rwi-essen.de

1. Introduction

Karl Popper’s hypothetico-deductivism is the prevailing empirical epistemology in economics and replication is at the core of his philosophy of science. His falsification claim is based on a “*reproducible effect which refutes the theory*” (Popper, 1959). According to Merton (1973, p. 276) it is replication and “*rigorous[ly] policing*” each other’s work that keeps scientists truthful and disinterested – often referred to as scientific self-correction. In this paper, we examine whether replications in economics are cited and whether replications change the citations of replicated papers. The self-correction narrative would arguably imply changing citation patterns in response to an unsuccessful replication.

The need for replications in economics has been debated over the past decades (Clemens, 2017; H. M. Collins, 1991; Dewald et al., 1986; Hamermesh, 2007; Leamer, 1983; Mirowski & Sklivas, 1991; Whaples, 2006). More recently, the profession has experienced noteworthy improvements in terms of preregistration and data sharing policies (Christensen & Miguel, 2018; Miguel, 2021), but replications are still rare (Ankel-Peters et al., 2023a).¹ At the same time, new meta-evidence indicates various forms of replicability problems and thereby emphasizes the need for more replications (Askarov et al., 2022; Brodeur et al., 2016, 2020; Camerer et al., 2016; Chang & Li, 2022; Dahal & Fiala, 2020; Ferraro & Shukla, 2020; Ioannidis et al., 2017; Peters et al., 2018; Vivalt, 2020).

We focus on replications published as comments in the *American Economic Review* (*AER*), one of the profession’s flagship journals. Comments in the *AER* challenge papers that were previously published in the journal, mostly based on robustness replications of that original paper (OP). The *AER* has always had a leadership role in the profession by publishing a relatively large number of comments and, more recently, by rigorously applying data sharing policies (AEA, 2023). We start by tracing

¹ Low replication rates have also been discussed in other disciplines such as psychology and medical science (Hensel, 2021; Maxwell et al., 2015; Open Science Collaboration, 2015).

the publication of comments over time and by showing results of a short survey we conducted among the *AER* editors since 1985.

In the main part of our paper, we ask whether a comment leads to a reappraisal of the OP's contribution to the literature. We investigate all comments published between 2010 and 2020 and find 56 comments, of which 37 received a reply by the original authors. We use Google Scholar citations to examine how often comments are cited and whether the citation trend of the replicated original paper has changed after the comment was published.

Our underlying assumption is that citations reflect the priors in the research community (Rubin & Rubin, 2021) and the community's appreciation of the publication (Bornmann & Daniel, 2008; Card & DellaVigna, 2020; Hamermesh, 2018; Siler et al., 2015). Therefore, provided the self-correction paradigm holds, a comment raising substantive concerns about an OP should either lead to a decrease in citations for the OP or the comment should be cited alongside it (Hardwicke et al., 2021). This part of our paper is similar to Coupé and Reed (2022) who examine the effect of 204 replications in economics on citation patterns of replicated papers, using econometric analysis. By focusing on replicated papers in the *AER*, we consider papers that are highly cited and have a strong influence on the research frontier (Teplitskiy et al., 2022). We thereby also check whether "*well-executed replications receive credit*" (Coffman et al. 2017).

Overall, we find that comments do not affect the OP's influence in the literature. More specifically, we find that comments are cited on average seven times per year since their publication – compared to on average 74 citations per year for the OP since publication of the comment. Comments are, hence, not cited much in absolute terms, and a lot less than the original paper. The latter implies that most OP citations ignore the comment. We furthermore find that the publication of a comment does not affect the OP's citation trend. To focus on influential high-quality citations, we zoom into how OPs are cited in the *Journal of Economic Literature* (JEL) and the *Journal of Economic*

Perspectives (JEP). We found 60 references in 53 *JEL/JEP* articles to 30 OPs from our sample. We corroborate our previous findings in that most OPs are cited in the *JEL* and *JEP* articles without mentioning the comment alongside it (only 18 out of 60 cite the comment).

In a next step, we address the fact that not all comments raise equally substantive concerns. Thus, the need for scientific self-correction differs across comments. Some comments are confirmative or qualify only parts of the OP, without questioning the OP's key contribution. Others fundamentally challenge the OP's main claims. This latter group should not be ignored by the literature if the self-correction paradigm is to hold. We read and rated all comments as to whether the respective comment, in our view, must be cited (against the alternative options 'sometimes' and 'never'). Moreover, we elicited the same assessment from the authors of original papers and of comments by means of a short author survey. We then test for the robustness of our results when focusing on those comments that are rated as must-cites or sometimes-cites and find that our interpretation holds even for these cases.

Our paper contributes to several literatures. We build on the growing meta-scientific literature in economics, which diagnoses increasing transparency standards (Christensen & Miguel, 2018; Miguel, 2021). Furthermore, we contribute to attempts in economics and other disciplines to shed more light on how replications are received in the scientific community (Coupé & Reed, 2022; Hardwicke et al., 2021; Schafmeister, 2021; Serra-Garcia & Gneezy, 2021; von Hippel, 2022), whether self-correcting mechanisms in science work (Ioannidis, 2012; Peterson & Panofsky, 2021; Vazire & Holcombe, 2021), and how the scientific community deals with demonstrably erroneous papers (Budd et al., 1998; Fernández et al., 2019). We complement Coupé and Reed (2022) by probing deeper into a smaller number of replications, those prominently published in the *AER*, while their work benefits from a larger and arguably more representative sample. Our sample, hence, probably represents the

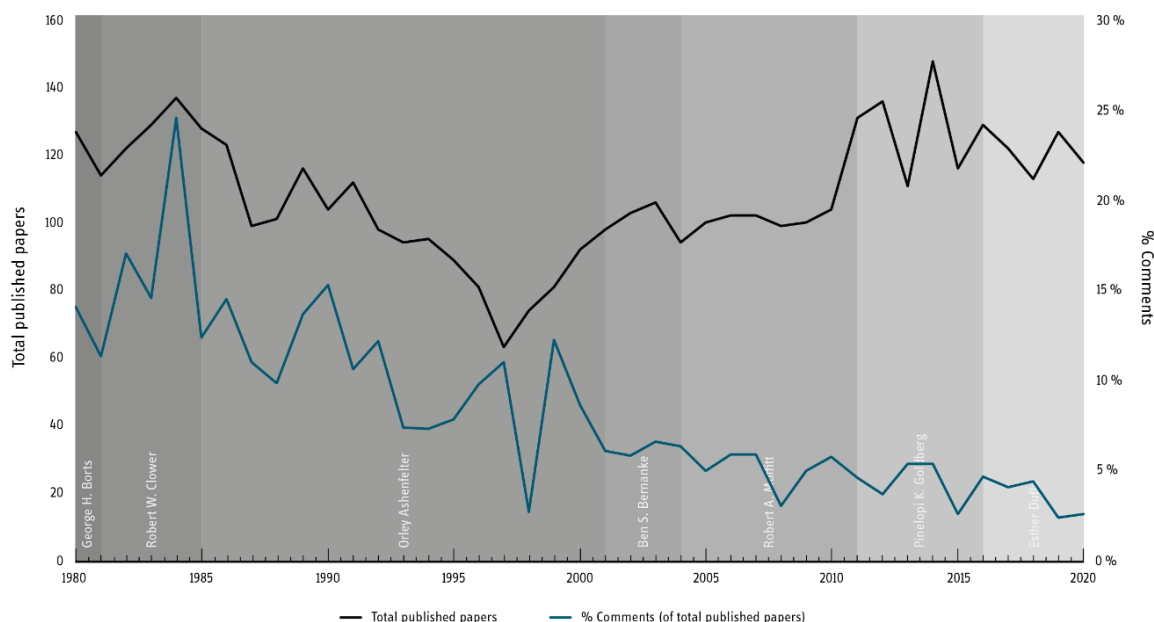
upper bound of citation prospects. It mutually strengthens our interpretation that both approaches lead to similar findings.

2. Replication and comments: The AER policy

The AER has a long history of publishing comments and its website contains a clear guideline for what comments are supposed to do and how they are handled (AEA, 2023):

“Comments submitted to the Review are refereed both by the author of the article being commented on and by other referees. Replies to Comments are sent to the author of the Comment and to other referees. There is no automatic right to Reply; the author of a Reply must provide substantive and material discussion of the issues in question. Comments and Replies which appear only on the AER web page are also sometimes considered.² These papers go through the same refereeing process as all Comments and Replies, but may be judged to be more appropriate for internet posting instead of publication in the printed AER.”

Figure 1: Papers and comments published in the AER between 1980 and 2020



Notes: Gray-shaded areas indicate the periods of AER editors-in-chief. Comments are elicited for each volume. Source: Own data.

² We checked the AER’s “archived internet comments” but only found one, published online in 2006 during Robert Moffitt’s term (accessed on July 11, 2022).

Throughout the 1980s and into the 1990s, more than 10% of the papers published in the *AER* were comments (Figure 1). This share has declined considerably over time, down to 2-3% in most recent years. Ankel-Peters et al. (2023) investigate this in more detail and find that in the 1980s most comments were theoretical. Since then, the share of empirical papers in the *AER* has increased considerably, which has not been accompanied by a simultaneous increase in empirical comments.

We conducted a short survey among the *AER*'s editors since 1985 to obtain a better understanding of the *AER* comment policy over time. Thankfully, all editors responded to our three-question survey, and all agree that there has not been an explicit change in the policy regarding the publication of comments. The following quotes are interesting attempts to explain the development over time:

Orley Ashenfelter:

The comment/reply format is a tedious one to referee. In addition, I believe that in tenure decisions comments do not receive the same 'points' as other publications. The change to judging publications on a point system probably started in the 1980s. This clearly reduces author incentives to write comments.

Ben Bernanke:

*It's a little surprising that there are fewer comments now because 12 *AER* issues plus four field journals mean a lot more available space. Maybe what used to be comments are now more likely to be expanded and accepted as regular papers. *Economic Insights*, another new journal, also publishes shorter papers.³*

Robert Moffitt:

I know that many people feel, today, that submitting a comment has the problem that the author will almost surely reply if the comment is negative, and that will generate many months of back-and-forth debate with the original authors. [...] I would not be surprised if many people also don't think a comment is particularly strong on a c.v. I suspect that many people feel it is just better to write a new, original paper which

³ Indeed, in 2019, the AEA launched a new journal, the *American Economic Review: Insights* (*AER:I*). However, the *AER:I*, has not published any comment since its inception. What is more, in a companion paper, we have searched for replications in the *AER* and other top journals and only found a small number that qualify as replications but are not published as comments (Ankel-Peters et al., 2023a).

explicitly or implicitly criticizes the original paper, than submit a comment.

Pinelopi Goldberg:

One hypothesis is that the published research has become both more complicated and more rigorous. There are many complaints about the increasing length of published papers [...]. Perhaps the flip side of longer papers is that they cover more ground, provide more robustness checks, and leave fewer open questions.

Esther Duflo:

We publish comments when the point made is of significant interest for the general readership, so either when results in very influential papers are overturned or there is a methodological contribution in the comment (and the comment is correct as far as we can see).

The *AER* is also a forerunner in terms of code and data sharing policies. It is important to demarcate the replication work that *AER* comments are based on (to the extent they are empirical, see Section 3.2) from the newly established replications conducted by the AEA data editor (Vilhuber, 2019). For this demarcation, we refer to the nomenclature on different replication sub-types defined by Dreber and Johannesson (2022) and the Institute for Replication (2022, see Table 1).⁴

Table 1: Replication definitions

Author(s), Year	Category	New paper uses the same...		
		Specification	Population	Sample
Institute for Replication; Dreber and Johannesson (2022)	Computational Reproduction	✓	✓	✓
	Recreate Reproduction ¹	✓/X ²	✓	✓
	Robustness Replication	X	✓	✓
	Direct Replication	✓	✓/X ³	X
	Conceptual Replication	X	✓/X ³	X

Notes: ¹Dreber and Johannesson (2022) introduce this additional category which differs from “computational reproduction” only in that it emphasizes the usage of raw data and not having the analysis code of the original paper. This category is not included in the *I4R* definition. ²The specification in the reproduction is not always identical to the original paper as the replicator does not have access to the original code but tries to recreate the analysis based on the given information in the original paper. ³*I4R*’s definitions of direct and conceptual replication only require new data but it does not matter if it is from the same population or not. Dreber and Johannesson (2022) further subdivide between the same, similar, and different populations.

⁴ Similar definitions with different nomenclatures exist, see for example Clemens (2017), Hamermesh (2007) and Freese and Peterson (2017). They all share very similar dimensions to distinguish different sub-types, that is, according to whether the replication uses the same specification, population, and sample. See also Ankel-Peters et al. (2023) for a more detailed review.

Already in 2005, during Robert Moffitt’s tenure, the AEA launched and implemented a new policy which made data sharing mandatory. Moreover, in 2018, they appointed a data editor to rigorously enforce the policy by conducting what Dreber and Johannesson (2022) call a *computational reproduction* on every accepted paper. The main purpose of this policy is to check whether the data and code are accessible and complete, and to ensure that the code reproduces the results (Vilhuber, 2019). Virtually all comments in our sample, if empirical, are based on *robustness replications*, *direct replications*, or *conceptual replications*. Perhaps closest to a *computational reproduction* are a few comments that find coding errors in the original study, but it is unlikely that the reproductions conducted as part of the AEA-checks would have uncovered the deeper coding issues that underlie these comments. Nevertheless, this rigorous data and code sharing policy might have signaling effects altering the incentive structure towards transparency and replicability (see e.g., Askarov et al., 2022).

3. Citation patterns for comments and original papers

3.1 Methodology

We conducted a systematic review covering eleven volumes of the *AER* from 2010 until 2020. We screened the *AER* website for all papers that included the word “comment” in the title. In total, we found 56 comments, written on 53 OPs⁵, 37 of which also received a reply from the original authors. For every OP, comment, and, if applicable, reply – henceforth a *debate* – we elicited the number of citations in Google Scholar (GS).⁶ We use the average annual citations since publication of the comment

⁵ Two OPs received more than one comment (Andreoni & Sprenger, 2012; Long & Ferrie, 2013). The comment by Rothstein (2017) replicates two papers: Chetty et al. (2014a) and Chetty et al. (2014b). After successfully replicating both papers, though, the comment focusses on the former. Thus, we only include Chetty et al. (2014a) in this list of OPs underlying our analyses.

⁶ This task was done between February 15 and 19, 2022. We also considered citations in Web of Science (WoS) but recent studies report a very high correlation between GS and WoS metrics – above 90% for economics (Hamermesh, 2018; Martín-Martín et al., 2018).

in the *AER* as the main citation indicator throughout the paper to ensure the comparability of citation counts between OP and comment.

3.2 Author and original paper characteristics

Table 2 shows some author characteristics to examine whether comment authors differ from original authors in terms of career status and influence. Authors of OPs are more senior and more influential than comment authors (as measured by Top5⁷ publications and GS citations). Some comment authors are likewise established researchers, but at the time of writing the comment many were at the beginning of their career. (See Tables C3-C5 in the Online Appendix for more author features; location, field of work, and highest obtained degree).

Table 2: Author characteristics

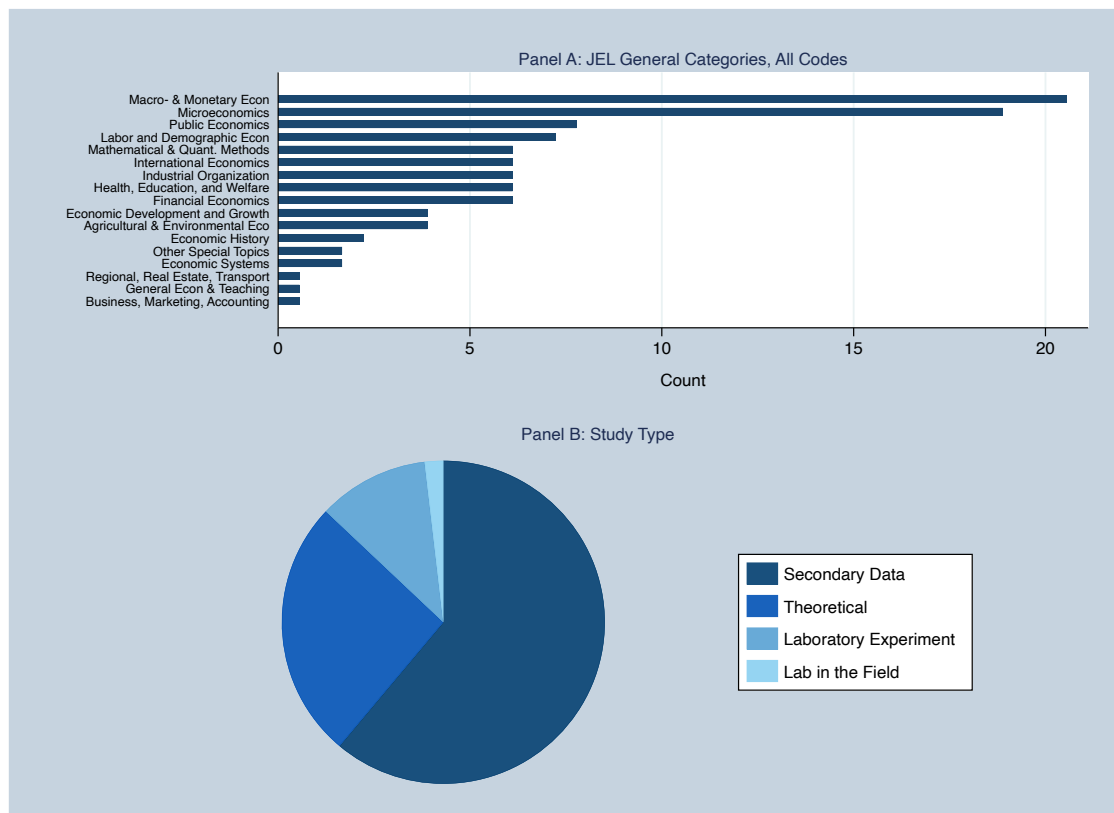
	OP authors		Comment authors	
N (Number of papers)	53		56	
N (Number of authors)	119		111	
Size of author team (average)	2.25		1.98	
	Team average	Maximum per team	Team average	Maximum per team
Number of years since PhD in t-1 ^{1,2}	15.1	20.0	9.0	13.6
# of Top 5 publications in t-1	5.9	10.3	1.4	2.25
Share of authors with <100 GS citations in t-1	0.23	0.49	0.63	0.84
GS citations in total	15,808	29,930	5,471	10,739
GS citations in 2021	1,409	2,675	463	855
GS citations in year before comment publication	839	1,492	277	542

Notes. ¹ t-1 is the year before comment publication. ² 97% of authors have a PhD. The number of observations is different for the indicator 'Average number of years since PhD' due to missing data (OP: 110, Comment: 108). GS = Google Scholar.

OPs cover a broad range of topics, but the *JEL*-code '*macro- and monetary economics*' dominates (Figure 2, Panel A). Most OPs use secondary data (33), the rest are either conceptual (i.e., theoretical or methodological, but definitely non-empirical) or laboratory experiments. Only one paper uses a laboratory experiment in the field and there is no Randomized Controlled Trial.

⁷ The Top5 journals in economics are: *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *The Quarterly Journal of Economics*, and *The Review of Economic Studies*.

Figure 2: JEL-codes and methods of original papers (OPs)



Notes. Panel A shows all JEL codes of all 53 OPs in our sample (N=180). Panel B is on a per-paper basis, i.e., N=53. Panel A is based on JEL codes, while panel B is based on our own judgment. Whittington et al. (1990) does not have any JEL codes and is, therefore, not included in panel A. Panel B includes one paper that collected primary but non-experimental data (Bonjour et al., 2003). We include it in the “Secondary Data” category. We coded all non-empirical papers as “Conceptual”.

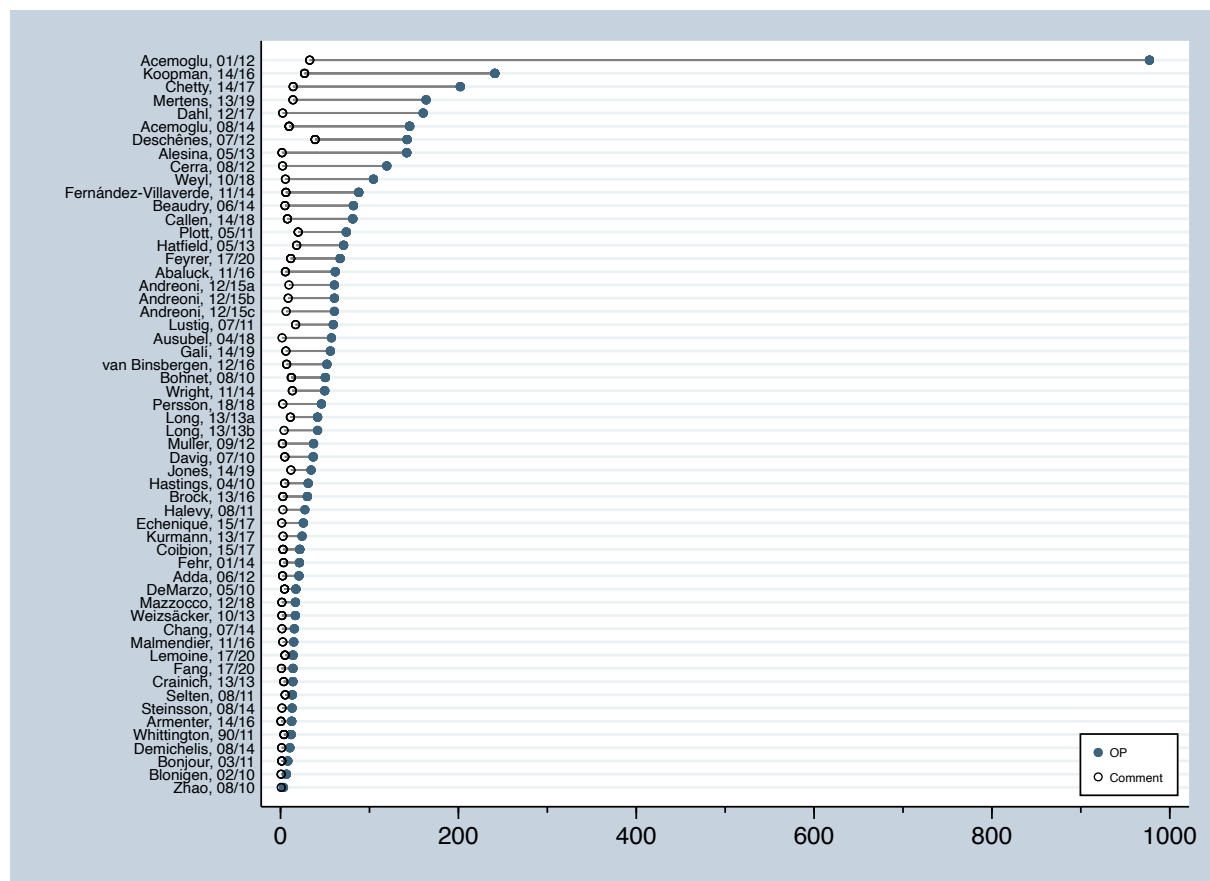
3.3 Are comments cited alongside the original paper?

Table A1 in the Appendix (at the end of this paper) provides comprehensive descriptive citation statistics for all OPs and comments in our sample, including total citations, average annual citations since the comment’s publication and the citation ratio of comment to OP. Figure 3 shows the average annual citations of the original papers (since publication of the comment) and their respective comments in our sample.⁸ The difference between citations of OPs and comments is large: Total average citations are 15 times higher for the OPs than for comments, but also the average annual citation count of OPs since comment publication is more than ten times higher (see Table A1). Most OPs are influential papers with total citation counts way above

⁸ Only four of the 56 comments have at least ten citations in total prior to their *AER* publication, probably because discussion paper versions had circulated before: Albouy (2012), Burnside (2011), Cheung (2015), and Fisher et al. (2012).

the *AER* average, but some OPs also have low citation counts. The debate between Acemoglu et al. (2001) and Albouy (2012) is a striking outlier: While the comment is the second most cited comment, the OP dwarfs its citation count.

Figure 3: Average annual citations – difference between original paper and comment

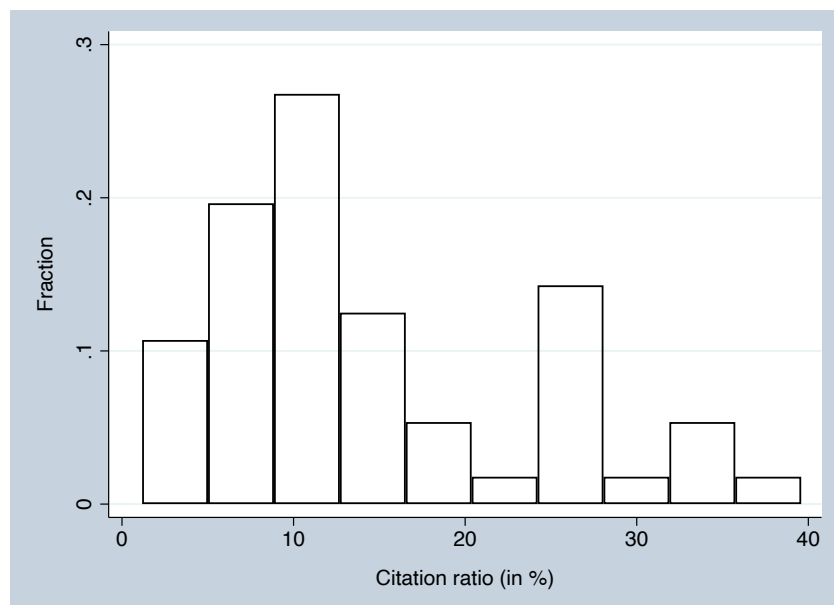


Notes. We included all comments published in the *AER* between 2010 and 2020 and their respective OP. Citations are counted since comment publication for each debate. OPs have blue markers; comment markers are dark gray. The labels on the y-axis show the first author of the OP, its publication year, and the publication year of the comment. The OP by Andreoni and Sprenger (2012) received three comments that we mark by the letters “a” (Miao & Zhong, 2015), “b” (Cheung, 2015), and “c” (Epper & Fehr-Duda, 2015). Similarly, the paper by Long and Ferrie (2013) received two comments that we mark with “a” (Xie & Killewald, 2013) and “b” (Hout & Guest, 2013).

The far-right column in Table A1 shows the ratio of all comment citations to OP citations. It is further visualized in Figure 4. This is a key indicator, because to maintain the claim that economics is self-correcting, one would expect a high ratio. The average citation ratio across all debates is at 14%. Looking at the distribution in Figure 4, 50% of the debates have citation ratios of less than 11%. No citation ratio is higher than 40% and only four of the 56 comments in our sample have an average annual citation ratio above 30%: Brunner et al. (2011) at 39.6%, Mattauch et al. (2020)

at 35.7%, Caselli and Ciccone (2019) at 34.0%, and Crump et al. (2011) at 32.3%. Hence, for most debates, the comment is largely ignored by the future literature.

Figure 4: Citation ratio distribution



Notes. Citations are counted since comment publication for each debate.

Hitherto, we have implicitly assumed that citations always express scientific appreciation, while, in principle, citations could also dismiss the referenced paper's content. To scrutinize this, we briefly investigate *how* the OPs are cited by using the scite.ai tool *Reference Check*, which classifies citation statements in referencing papers into 'supporting', 'mentioning' or 'contrasting' the referenced paper (with 'unclassified' as a fourth option in case the tool cannot assign the statement to one of the three categories). We find that almost all citation statements about OPs in our sample (93.8%) are categorized as 'mentioning'. Further, 4.6% are 'supporting', only 0.6% are 'contrasting', and 1% are 'unclassified'.

3.4. Citations in the *Journal of Economic Literature* and *Journal of Economic Perspectives*

We now zoom into citations of particularly high influence, namely those in the *JEL* and the *JEP*. These articles arguably represent the highest standard in economics, and one can assume that they are written with exceptional care and expertise. As we will

show, the citation ratio in *JEL/JEP* articles is higher than the ratio in the previous section, but nevertheless the majority of such high-quality citations ignore the comment.

We searched Web of Science for all *JEL* and *JEP* articles (excluding book reviews and eulogies) that cite our sample of OPs and comments and found 53 individual papers, 36 from the *JEL* and 17 from the *JEP*, which cite 30 of our debates.⁹ Some of these 53 *JEL/JEP* articles cite more than one OP, which is why we have 60 references in our dataset.

Figure 5: Share of *JEL/JEP* papers citing either the OP or the comment, or both

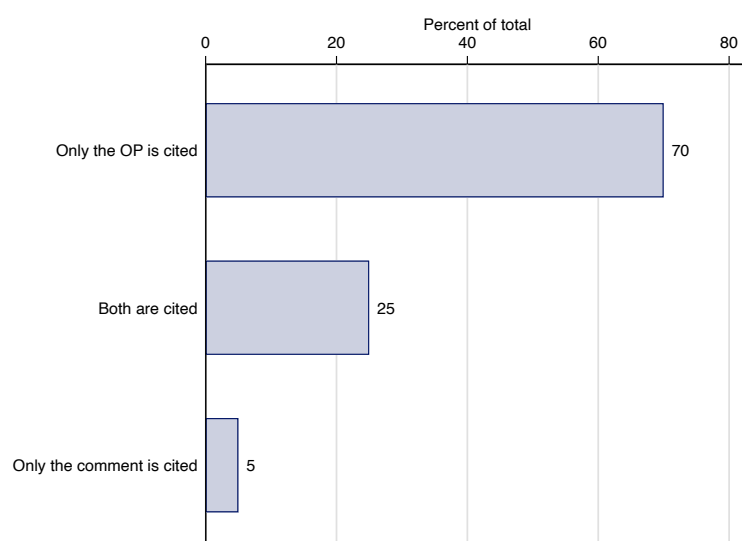


Figure 5 shows that 70% (or 42) of these references cite the OP without mentioning the comment. 25% (or 15) cite both and 5% (or 3) cite only the comment. Dividing the number of comment citations (18) by the number of OP citations (57), results in a *JEL/JEP*-exclusive citation ratio of 32% (versus 14% in Section 3.3 and in Table A1).

Next, in Table 3 we take a closer look at the quality of citations by examining how exactly the OPs and comments are cited in the *JEL/JEP* articles.¹⁰ As opposed to Figure 5, where a cited OP is only counted once even if it appears multiple times in the same

⁹ A few OPs are cited in multiple *JEL/JEP* articles; Acemoglu et al. (2001), for example, is cited in 12 *JEL/JEP* articles.

¹⁰ We also disaggregate the papers for each journal and show the results in Table E7 in the Online Appendix.

JEL/JEP paper, we now refer to each appearance as one ‘citation statement’. Table 3 counts these citation statements individually and rates them as either ‘regular’ or ‘prominent’ (definitions can be found in the Table’s note). We find that most citation statements (36, or 56%) mention the OP ‘prominently’, without mention of the comment. Another 18 (or 28%) cite the OP in a ‘regular’ way, again with no mention of the comment (see column 1). Those *JEL/JEP* articles that do mention the comment (column 2 and 3) still dedicate more attention to the OP than to the comment. Only one *JEL/JEP* article citing the comment summarizes the debate in the main text. In five cases, a *JEL/JEP* article cites both papers and mentions that there is a debate but does not summarize its content.¹¹

Table 3: Citation statements to original papers (OPs) and comments in the *Journal of Economic Literature* and the *Journal of Economic Perspectives*

Categories of OP citations	OP citation statements...		Comment citation statements
	...if only the OP is cited (1)	...if the OP and the comment are cited (2)	(3)
Main text, prominent	36	19	1
Main text, regular	18	10	6
Main text, mention of the debate	NA	5	5
Main text, summary of the debate	NA	1	1
Footnote	7	2	6
Only in reference list, not in text	3	0	0
Sum	64	37	19
Total citation statements	120		

Notes. We exclude book reviews, eulogies, and the three references to the comment only (see Figure 5, lower bar). A "regular" citation statement is a citation in brackets after a general sentence, next to other references. A "prominent" citation statement means that at least one sentence is dedicated to the cited paper and it is mentioned explicitly. NA = Not applicable in cases where the category requires a comment citation but it is not cited.

3.5 Do comments affect citation trends of original papers?

Next, we investigate whether the publication of a comment leads to a reappraisal of the OP’s influence in the literature, measured by the citation count of the OP. A transparent way to do this, given the limited number of observations in our sample,

¹¹ Replies are only mentioned in seven out of 46 references.

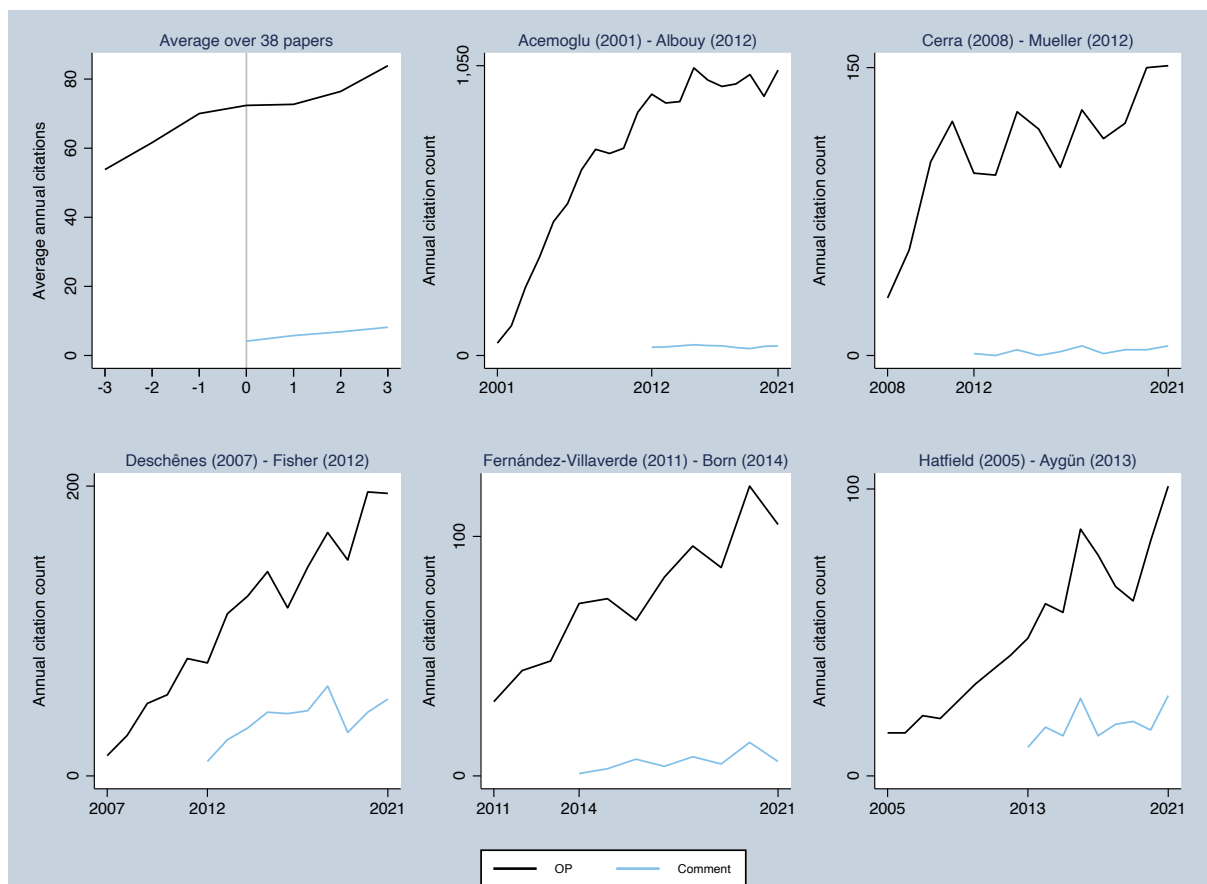
is via visual inspection of citation trends.¹² As we will show, the publication of comments does not lead to a decrease of OP citations. What is more, most OPs (46 or 82%) have higher average annual citation counts after the comment was published than before. The first panel of Figure 6 shows the average annual citations of OPs and comments in the three years before and after the comment publication (t_0): The positive trend of rising citations for the OP continues after t_0 . We test whether this observation holds for different time horizons before and after t_0 , and it does (see Figure D1 in the Online Appendix). The caveat of this robustness test is that the sample size of eligible comments decreases quite considerably the wider the sample period is, i.e., the more years before and after t_0 we include.

It is likely that some comments had circulated as discussion papers before they were published in the *AER*. Yet, while 30 of 56 comments have been cited prior to their publication in the *AER* (probably as discussion paper versions), only four received more than 10 citations. Moreover, given that most OPs' annual citation count increases over time, it is very unlikely that the discussion paper versions had a strong effect on OP citations.

We furthermore show individual citation trends before and after comment publication for a selection of the debates in the remaining panels of Figure 6 (see Figures D3-D6 in the Online Appendix for citation trends of all OPs and comments). There are few cases among all debates where the citation trend of the OP is not as steep or even stagnates post comment publication – and we cannot rule out that this is due to the comment. However, even in these cases, the OP's citation count trend flattens on a considerable high level and annual citations do not decrease. Overall, these trends confirm our verdict for the average across all OPs: the comment's publication does not lead to a reappraisal of the OP's influence in the literature.

¹² See Coupé and Reed (2022) for an econometric analysis of a larger sample of replications and original papers.

Figure 6: Original Papers' and comments' average annual citation counts before and after comment publication



Notes. In the first panel, 0 is the year of comment publication in the AER. For each panel, the black line depicts the annual citation count of the original paper, and the light blue line depicts that of the comment. The first paper in the title is the original paper, and the paper mentioned second is the comment. In both cases, we only depict the first author for space reasons.

Our visual analysis does not compare the citation trend of OPs in our sample to counterfactual trends. It is indeed not our intention to make a precise causal statement of how much a comment affects the OP's citation trend. Yet, we might miss relative decreases of OP citation counts vis-à-vis similar non-replicated papers that are induced by the comments' publication. To probe a bit into this, we approximate the OP's counterfactual trend by plotting the citation counts for all *AER* papers published in the same issues as the 20 most cited OPs in our sample (see Figures D7 and D8 in the Online Appendix). We find that the citation trends of the OPs follow a similar

pattern compared to their respective same-issue-papers, reassuring our above reading of OP citation trends.¹³

4. Subjective Ratings and Author survey

4.1 Subjective ratings

Not all comments put forward equally fundamental criticism of the OP. Comments diagnosing (or claiming to diagnose) deeper problems are arguably more important to be cited. For that reason, we assess the substance of each comment in this section. We first rated each debate ourselves and, second, we conducted a survey with the authors of OPs and comments to obtain their assessment. For our own subjective rating, all three co-authors read the entire debate and answered the question “Should the comment be cited whenever the OP is cited?”, with three possible answers: a) Yes, in virtually all cases; b) Yes, but only in some cases, c) No, the comment does not have to be cited. In case our ratings deviated, disagreements were resolved through discussion. We asked the authors the same question in our survey.

The subjective leeway in rating the debates is obvious. For example, OPs might include several results of equal importance, but the comment only criticizes one. The reply might acknowledge problems with this one result but claims that the OP provided several results, and the others still hold. Other, similar, scenarios are possible. Thus, any replication needs to be qualified and a debate is inevitable. Some comments stake out the extent of their criticism and its implication very clearly, others do not. Most replies, in turn, do not acknowledge the diagnosed problem, or they acknowledge parts of it, but question the extent.

¹³ See also Coupé and Reed (2022) for matching-based comparison of confirmatory and contradictory replications and their effect on citations of replicated papers.

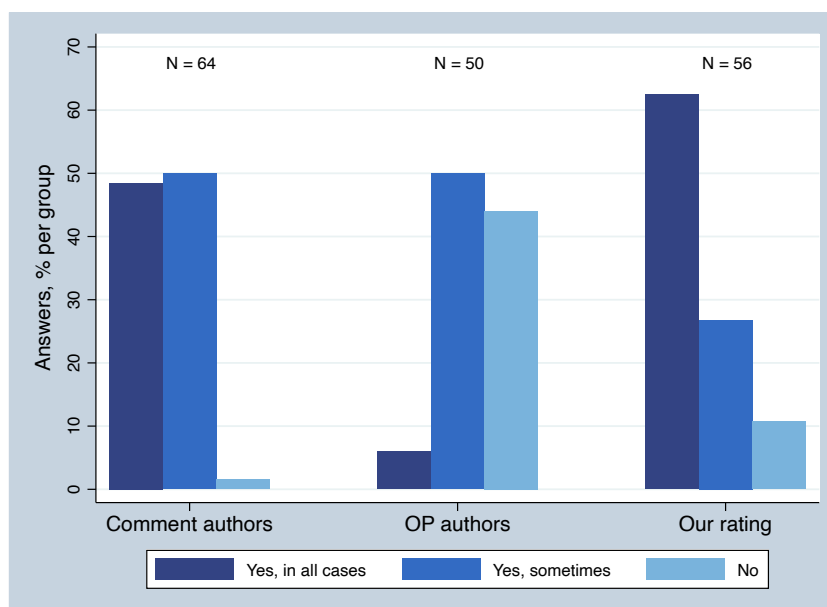
We also used the author survey to find out more about the debate.¹⁴ We programmed the survey in Qualtrics and sent out the link to all authors via e-mail.¹⁵ That is, if a paper is written by three authors, all three received a survey link and were informed that their co-authors were also contacted. We assured respondents anonymity in our invitation e-mail and, therefore, will not report article-specific responses in this paper. Response rates are shown in Table 4.

Table 4: Author survey response rates

Survey	By authors			By papers		
	Contacted	Replied	%	Contacted	Replied	%
Comment	98	66	67%	51	43	84%
OP	111	53	48%	51	38	75%
Total	209	119	57%	102	81	79%

Notes. The response rates by papers counts a paper where at least one author responded to our survey as “replied”. We depict the response rates separate for the four groups of authors: 1) Authors of comments that did not receive a reply from the original authors, 2) Authors of comments that did receive a reply from the original authors, 3) Authors of original papers that did not write a reply, and 4) Authors of original papers that did write a reply. We did not contact authors of debates where the OP received more than one comment. This applies to two OPs: Andreoni and Sprenger (2012) and Long and Ferrie (2013). In addition, 3 authors deceased, and we could not find working e-mail addresses of 5 authors.

Figure 7: Ratings: “If a researcher cites the original paper, do you think the comment should be cited, as well?”



Notes. The number of observations reflect the responses we received from the OP and comment authors to this question, and in our case our rating of each of the 56 comments.

¹⁴ All details on the implementation of the survey, the comprehensive list of questions, the scripts of the survey, invitation e-mails, and reminders, as well as the details on the feedback rates can be found in Online Appendix B.

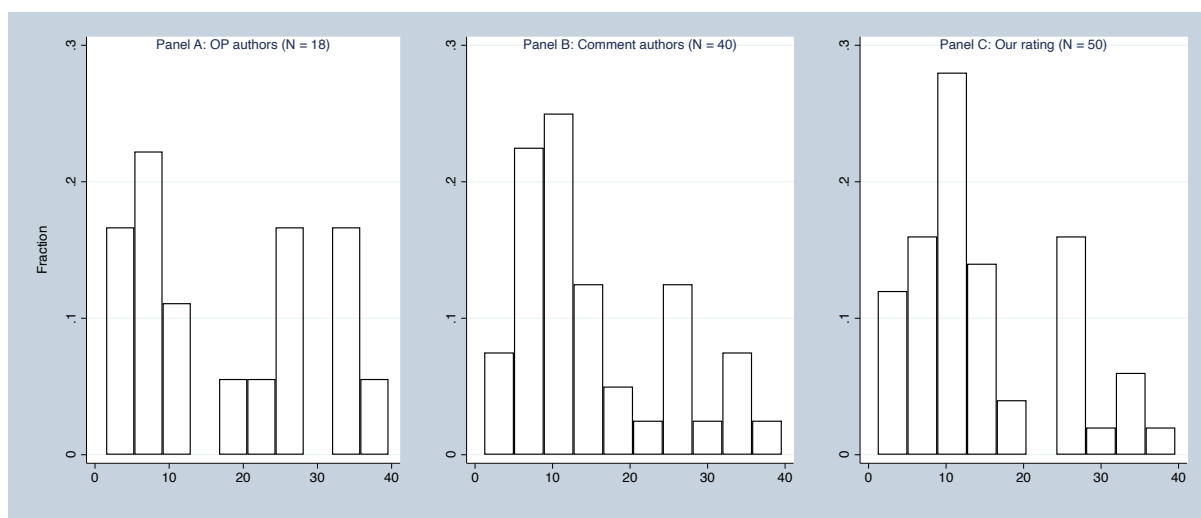
¹⁵ We exclude the debates where the OP received multiple comments: Andreoni and Sprenger (2012) and Long and Ferrie (2013).

Figure 7 shows the results for the author group ratings and our own. The difference between comment authors and original authors is striking, while our own rating is closer to the comment authors.

4.2. Citation patterns for comments rated as ‘must-cite’ and ‘sometimes-cite’

We use the ratings by the three author groups to check whether our previous findings change if we exclude those comments that are rated as ‘never-cite’. We first corroborate the distribution analysis of the citation ratio, now only for those comments that were rated either as ‘must-cites’ or ‘sometimes-cites’. Figure 8 confirms our previous analysis in Figure 4 for comment authors and our rating.

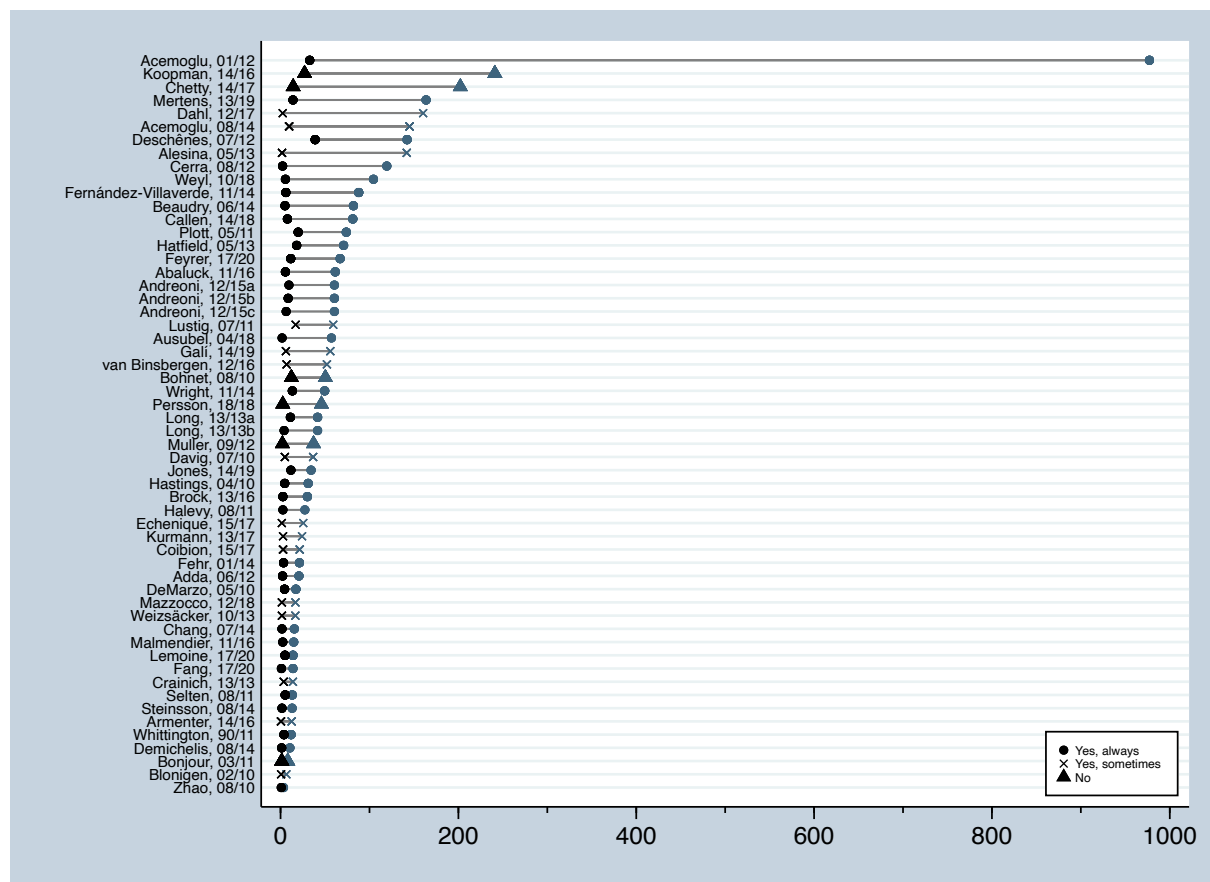
Figure 8: Distribution of the citation ratio (in %) for debates with comments rated as ‘must-cite’ and ‘sometimes-cite’



Notes. The number of observations reflect the number of debates included in each panel, i.e., all debates where the comment was scored as either “yes, [should be cited in all cases]” and “sometimes”. For the few debates where we received multiple responses with different scores from the same author group (original authors or comment authors), we calculated the average score. We only included debates with a score equal or less than 2. Our scoring system was coded as 1 - “yes, [should be cited in all cases]”, 2 - “sometimes”, and 3 - “never”.

Panel A with the OP authors suffers from a much smaller sample size, but it does contain the few comments at the upper bound of our citation ratio range. Regardless, even for those comments rated as ‘must-’ or ‘sometimes-cites’ by the OP authors, most reveal citation ratios around or below 10%.

Figure 9: Average annual citations - difference between original paper and comment (by must-cite/sometimes-cite/never-cite)



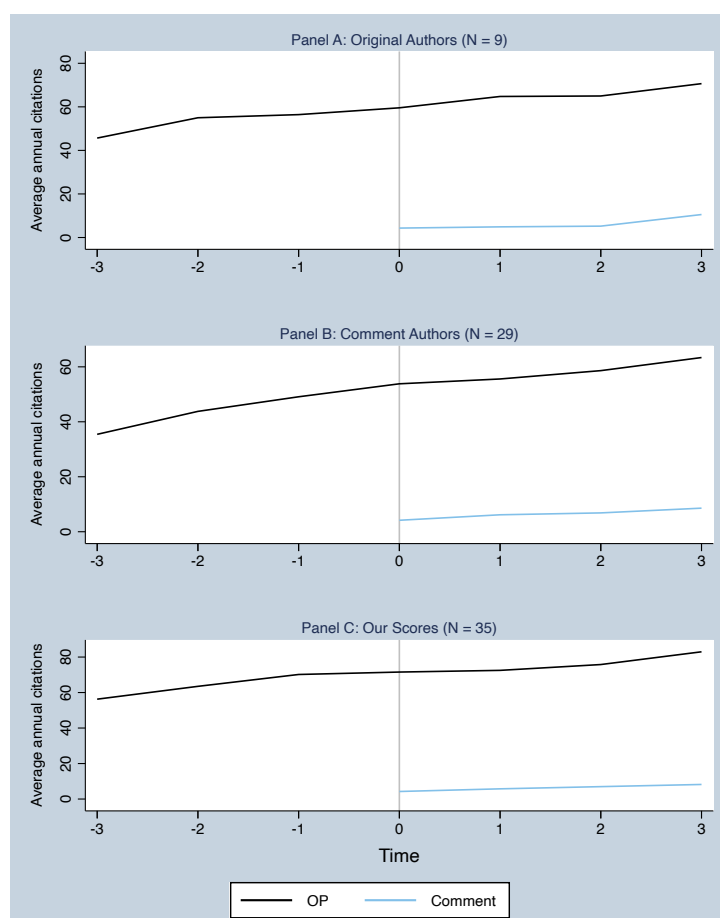
Notes. We included all comments published in the *AER* between 2010 and 2020 and their respective OP. Citations are counted since comment publication for each debate. OPs have blue markers; comment markers are black. The coding of the debates, i.e., the marker shape, is based on our subjective ratings of the debates as discussed in Section 3. The labels on the y-axis show the first author of the OP, its publication year, and the publication year of the comment. The OP by Andreoni and Sprenger (2012) received three comments that we mark by the letters “a” (Miao & Zhong, 2015), “b” (Cheung, 2015), and “c” (Epper & Fehr-Duda, 2015). Similarly, the paper by Long and Ferrie (2013) received two comments that we mark with “a” (Xie & Killewald, 2013) and “b” (Hout & Guest, 2013).

Figure 9 shows annual citations for comments and OPs, now distinguished by our own rating. Comments come only close to OPs for debates in which the OP itself is hardly cited. Several heavily cited OPs received comments that we rated as ‘must-cites’ – and yet the comments are barely cited. It is noteworthy that some OPs at the bottom of the figure are hardly cited, and, at the same time, we rated them as ‘must-cites’.¹⁶ It is arguably possible that the comment in these cases has worked as a self-correcting mechanism.

¹⁶ In fact, in ten cases, the comment was published in the same year as the OP (four times) or only two years after (six times). Four out of the ten OPs have average annual citation counts below 25 since publication of the comment and are either rated as “must-cites” or “sometimes-cites”.

Figure 10 tests for the robustness of our citation trend analysis pre- and post-publication of the comment in Panel A of Figure 6 (see Figure D2 in the Online Appendix for additional robustness tests on different pre- and post-comment horizons and for paper specific citation trends see Figures D3-D6 in the Online Appendix). Again, our verdict holds: comments do not seem to lead to a reappraisal of the OPs' influence in the literature, even for 'must-cite' comments.

Figure 10: Average annual citation counts before and after comment publication for debates with comments rated as 'must-cite' or 'sometimes-cite'

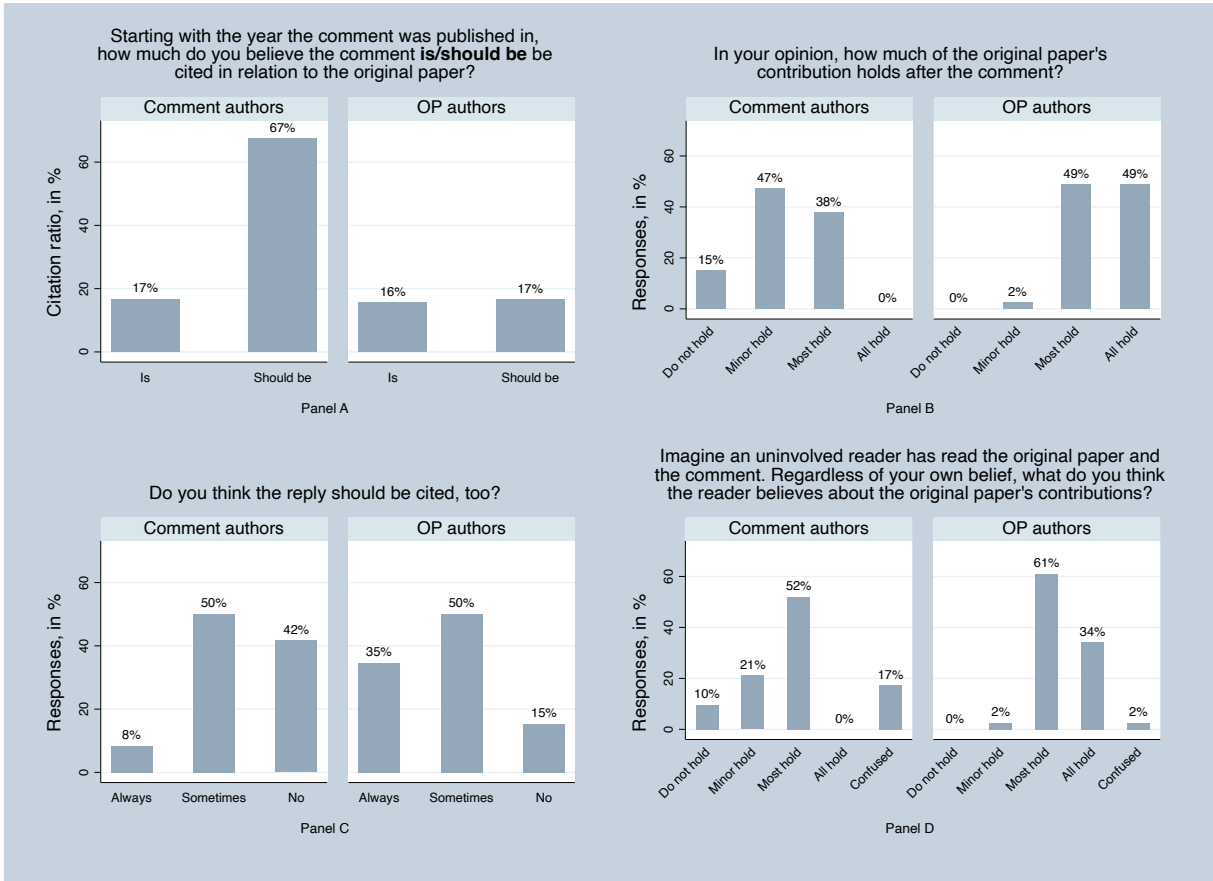


4.3. Author survey results

Some of the debates seem heavily controversial and, especially those with a reply by the original authors, are not easy to rate. Even for comments that in our opinion clearly change fundamental parts of the OP's results, replies push back harshly or at least try to maintain the OP's contribution. We therefore added a few questions to the author

survey on how the authors perceive the debate – and show some of the results in this section (see Online Appendix B for the comprehensive list of questions).¹⁷

Figure 11: Authors’ responses on the OP’s contribution



Notes. Values above the bars are rounded to the nearest integer. The exact answer options for Panels B and C were 1) “Do not hold anymore”, 2) “Only minor contributions hold”, 3) “Most important contributions hold”, 4) “Hold in their entirety”, 5) “The reader is likely confused”. Respondents could also choose “Don’t know” and “Refuse to answer”. For Panel B, three comment authors chose “Refuse to answer”, and for Panel D, two comment authors chose “Refuse to answer”, and two comment authors as well as one original author chose “Don’t know”. Because we omit responses, the percentages in Panels B, C, and D add up to 100% for each group of authors, except for OP authors in Panel D due to rounding.

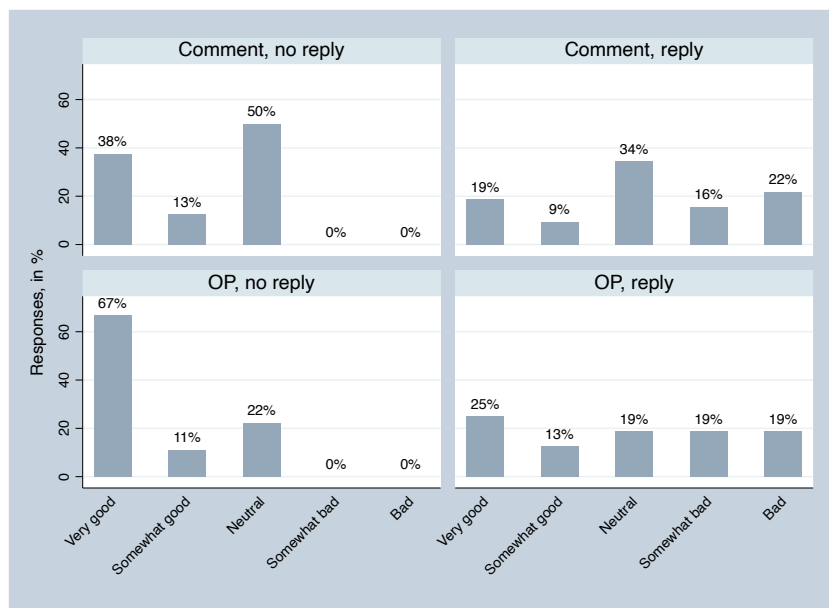
The panels in Figure 11 provide different perspectives on how much both groups diverge, sometimes drastically, also for their opinion on the contribution of the OP after the comment What we derive from this is that most uninvolved readers of the debates will have difficulties to reassess the OP’s influence in the literature.

One of the most frequently mentioned impediments to replications is that they are perceived as hostile. We asked all authors about their sentiment towards the interaction with the other author team and, as Figure 12 shows, the picture is mixed.

¹⁷ The results to all questions can be obtained from the authors upon request.

The demarcation line is whether original authors replied to the comment: for those debates without a reply, both author teams overwhelmingly think positively or neutral of the interaction. For those with a reply, in both teams a considerable percentage is unhappy with the interaction.

Figure 12: “How would you rate your interaction with the authors of the other paper?”



Note. Values above the bars are rounded to the nearest integer. Respondents could also choose “Don’t know” and “Refuse to answer”. One comment author chose “Refuse to answer” and one original author chose “Don’t know”. Because we omit these responses, the percentages add up to 100% for each group of authors. Differences are due to rounding. The numbers of observations are 16 for ‘Comment, no reply’, 32 for ‘Comment, reply’, 9 for ‘OP, no reply’, and 15 for ‘OP, reply’.

The unpleasant experiences in their interaction with the other team is not only visible in the numbers but also in the text responses for open questions we asked. For example, one comment author said: *“We were aiming to clarify their views on some key issues, but we found it difficult to get them to engage substantively”*. Another said that the relationship was good at first, but *“relations became more difficult when the nature of our criticism became clear”*. Some comment authors were clearly frustrated by the exchange, as this testament shows: *“We could not pinpoint all of the issues, because they would not share their data. Once the data was published, we could begin the difficult task of unpacking their (many) errors. One of the authors was a bit more receptive, but the more powerful of the two [...] simply wasn’t willing to engage substantively. It was not pretty.”* However, while we received much negative feedback from comment authors, original authors were

less talkative but one original author said: *“The authors sent many different results some in support of our original paper, some in conflict, but the comment they wrote only contained the most negative one, without mention that it was not robust. They were essentially trying to score some cheap points to get a publication.”*

5. Conclusion

In this paper, we have shown that replications, published as comments in the *AER*, are not cited much and have no discernable impact on the citation trends of the replicated original paper. We interpret this as evidence for the absence of self-correction mechanisms in economics. In doing so, we apply a narrow definition of scientific self-correction, what Peterson and Panofsky (2021) call ‘formal self-correction’. Formal self-correction relies on ‘diagnostic replication’¹⁸ and *“its outcome is some change to the original study that either emends or retracts it”*. To specify our interpretation, we hence provide strong evidence that economics is not subject to *formal* self-correction.

Our results are not at odds, though, with a broader definition, what Peterson and Panofsky (2021) call ‘organic self-correction’. This happens *“largely through the unpublished backchannels of a field. [...] Formal self-correction remembers wrongness; organic self-correction forgets that which is not useful.”* In fact, some of the barely cited OPs at the bottom of Figures 3 and 9 might have been subject to organic self-correction, catalyzed by the comment that got published shortly after (and hence, the literature has not even started to cite the OPs). Yet, several highly cited OPs in the upper part of those figures suggest anecdotally that economics is not very effective in organically self-correcting either.

Self-correction in economics is perhaps also difficult because the results of many replications are very controversial (see as well Ozier, 2021, and Roodman and

¹⁸ In a companion paper we introduce a term, policing replication, that is similar in meaning to what Peterson and Panofsky (2021) call ‘diagnostic replication’ (see Ankel-Peters et al., 2023).

Morduch, 2014). This holds true for our sample: for cases with a reply by the original authors, the extent to which a comment changes the contribution of the original paper is a matter of a fierce debate. For a neutral reader – like us – it is often confusing. It also resonates with results from the author survey we conducted, as well as with statements we obtained from the former *AER* editors. It also is similar to what Harry Collins called the ‘experimenter’s regress’: *“the problem is that, since experimentation is a matter of skillful practice, it can never be clear whether a second experiment has been done sufficiently well to count as a check on the results of the first”* (Collins, 1992, p. 2). We believe this problem is particularly hard to overcome in economics and other social sciences – disciplines that mostly work outside the laboratory – where the leeway for both researchers and replicators is very high (see Ankel-Peters et al., 2023; Breznau et al., 2022; Bryan et al., 2019; Huntington-Klein et al., 2021).

The absence of a clearcut definition of robustness and replicability raises questions about the extent to which empirical economics can live up to the Popperian definition of ‘science’. It does not have to, there are other reasonable epistemologies that are not falsified by the absence of replicability, like Imre Lakatos's understanding of scientific progress through the progressiveness of research programs. This would imply, though, a humbler interpretation of research results and more modest communication to the outside world. Irrespective of this deeper epistemological question, economics could do more to reveal its appreciation for replication. The *AER*, to begin with, deserves to be applauded for systematically publishing comments and a rigorous data sharing policy. It is yet surprising that the journal does not include links to the comments on the original papers’ website – something that is standard in other professions.

References

- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review*, 91(5), 1369–1401.
- AEA. (2023). *AER: Editorial Policy*. <https://www.aeaweb.org/journals/aer/about-aer/editorial-policy>
- Albouy, D. Y. (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *American Economic Review*, 102(6), 3059–3076. <https://doi.org/10.1257/aer.102.6.3059>
- Andreoni, J., & Sprenger, C. (2012). Risk Preferences Are Not Time Preferences. *American Economic Review*, 102(7), 3357–3376. <https://doi.org/10.1257/aer.102.7.3357>
- Ankel-Peters, J., Fiala, N., & Neubauer, F. (2023a). Do Economists Replicate? *I4R Discussion Paper Series No. 13*. <https://www.econstor.eu/handle/10419/267931>
- Ankel-Peters, J., Vance, C., & Bensch, G. (2023b). Spotlight on researcher decisions—Infrastructure evaluation, instrumental variables, and first-stage specification screening. *Ruhr Economic Papers*, No. 991.
- Askarov, Z., Doucouliagos, A., Doucouliagos, H., & Stanley, T. D. (2022). The Significance of Data-Sharing Policy. *Journal of the European Economic Association*, jvac053. <https://doi.org/10.1093/jeea/jvac053>
- Bonjour, D., Cherkas, L. F., Haskel, J. E., Hawkes, D. D., & Spector, T. D. (2003). Returns to Education: Evidence from U.K. Twins. *American Economic Review*, 93(5), 1799–1812. <https://doi.org/10.1257/00028280322655554>
- Bornmann, L., & Daniel, H. (2008). What Do Citation Counts Measure? A Review of Studies on Citing Behavior. *Journal of Documentation*, 64(1), 45–80. <https://doi.org/10.1108/00220410810844150>
- Breznau, N., Rinke, E. M., Wuttke, A., & Nguyen, H. H. V. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *PNAS*, 119(44). <https://doi.org/10.1073/pnas.2203150119>
- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics. *American Economic Review*, 110(11), 3634–3660. <https://doi.org/10.1257/aer.20190687>
- Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics*, 8(1), 1–32. <https://doi.org/10.1257/app.20150044>
- Brunner, C., Camerer, C. F., & Goeree, J. K. (2011). Stationary Concepts for Experimental 2 × 2 Games: Comment. *American Economic Review*, 101(2), 1029–1040. <https://doi.org/10.1257/aer.101.2.1029>
- Bryan, C. J., Yeager, D. S., & O'Brien, J. M. (2019). Replicator Degrees of Freedom Allow Publication of Misleading Failures to Replicate. *Proceedings of the National Academy of Sciences*, 116(51), 25535–25545. <https://doi.org/10.1073/pnas.1910951116>

- Budd, J. M., Sievert, M., & Schultz, T. R. (1998). Phenomena of Retraction: Reasons for Retraction and Citations to the Publications. *JAMA*, 280(3), 296. <https://doi.org/10.1001/jama.280.3.296>
- Burnside, C. (2011). The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk: Comment. *American Economic Review*, 101(7), 3456–3476. <https://doi.org/10.1257/aer.101.7.3456>
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating Replicability of Laboratory Experiments in Economics. *Science*, 351(6280), 1433–1436. <https://doi.org/10.1126/science.aaf0918>
- Card, D., & DellaVigna, S. (2020). What Do Editors Maximize? Evidence from Four Economics Journals. *The Review of Economics and Statistics*, 102(1), 195–217. https://doi.org/10.1162/rest_a_00839
- Caselli, F., & Ciccone, A. (2019). The Human Capital Stock: A Generalized Approach: Comment. *American Economic Review*, 109(3), 1155–1174. <https://doi.org/10.1257/aer.20171787>
- Chang, A. C., & Li, P. (2022). Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say “Often Not.” *Critical Finance Review*, 11. <https://doi.org/10.1561/104.000000053>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9), 2593–2632. <https://doi.org/10.1257/aer.104.9.2593>
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>
- Cheung, S. L. (2015). Comment on “Risk Preferences Are Not Time Preferences”: On the Elicitation of Time Preference under Conditions of Risk. *American Economic Review*, 105(7), 2242–2260. <https://doi.org/10.1257/aer.20120946>
- Christensen, G., & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *Journal of Economic Literature*, 56(3), 920–980. <https://doi.org/10.1257/jel.20171350>
- Clemens, M. A. (2017). The Meaning of Failed Replications: A Review and Proposal. *Journal of Economic Surveys*, 31(1), 326–342. <https://doi.org/10.1111/joes.12139>
- Collins, H. (1992). *Changing order: Replication and induction in scientific practice*. University of Chicago Press.
- Collins, H. M. (1991). The Meaning of Replication and the Science of Economics. *History of Political Economy*, 23(1), 123–142. <https://doi.org/10.1215/00182702-23-1-123>
- Coupé, T., & Reed, W. R. (2022). Do Negative Replications Affect Citations? *University of Canterbury, Department of Economics and Finance Working Papers in Economics* (No. 22/16).

- Crump, R., Shah Goda, G., & Mumford, K. J. (2011). Fertility and the Personal Exemption: Comment. *American Economic Review*, 101(4), 1616–1628. <https://doi.org/10.1257/aer.101.4.1616>
- Dahal, M., & Fiala, N. (2020). What Do We Know about the Impact of Microfinance? The Problems of Power and Precision. *World Development*, 128, 104773. <https://doi.org/10.1016/j.worlddev.2019.104773>
- Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *The American Economic Review*, 76(4), 587–603.
- Dreber, A., & Johannesson, M. (2022). *A Framework for Evaluating Reproducibility and Replicability in the Social Sciences*. Mimeo.
- Epper, T., & Fehr-Duda, H. (2015). Comment on “Risk Preferences Are Not Time Preferences”: Balancing on a Budget Line. *American Economic Review*, 105(7), 2261–2271. <https://doi.org/10.1257/aer.20130420>
- Fernández, L. M., Hardwicke, T. E., & Vadillo, M. A. (2019). *Retracted Papers Clinging on to Life: An Observational Study of Post-Retraction Citations in Psychology* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/cszpy>
- Ferraro, P. J., & Shukla, P. (2020). Feature—Is a Replicability Crisis on the Horizon for Environmental and Resource Economics? *Review of Environmental Economics and Policy*, 14(2), 339–351. <https://doi.org/10.1093/reep/reaa011>
- Fisher, A. C., Hanemann, W. M., Roberts, M. J., & Schlenker, W. (2012). The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment. *American Economic Review*, 102(7), 3749–3760. <https://doi.org/10.1257/aer.102.7.3749>
- Hamermesh, D. S. (2007). Viewpoint: Replication in Economics. *Canadian Journal of Economics/Revue Canadienne d'économique*, 40(3), 715–733. <https://doi.org/10.1111/j.1365-2966.2007.00428.x>
- Hamermesh, D. S. (2018). Citations In Economics: Measurement, Uses, and Impacts. *Journal of Economic Literature*, 56(1), 115–156. <https://doi.org/10.1257/jel.20161326>
- Hardwicke, T. E., Szűcs, D., Thibault, R. T., Crüwell, S., van den Akker, O. R., Nuijten, M. B., & Ioannidis, J. P. A. (2021). Citation Patterns Following a Strongly Contradictory Replication Result: Four Case Studies From Psychology. *Advances in Methods and Practices in Psychological Science*, 4(3), 251524592110408. <https://doi.org/10.1177/25152459211040837>
- Hensel, P. G. (2021). Reproducibility and Replicability Crisis: How Management Compares to Psychology and Economics – A Systematic Review of Literature. *European Management Journal*, 39(5), 577–594. <https://doi.org/10.1016/j.emj.2021.01.002>
- Hout, M., & Guest, A. M. (2013). Intergenerational Occupational Mobility in Great Britain and the United States Since 1850: Comment. *American Economic Review*, 103(5), 2021–2040. <https://doi.org/10.1257/aer.103.5.2021>

- Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J. R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., Saavedra, M., & Stopnitzky, Y. (2021). The Influence of Hidden Researcher Decisions in Applied Microeconomics. *Economic Inquiry*, 59(3), 944–960. <https://doi.org/10.1111/ecin.12992>
- Institute for Replication. (2022). *Defintions*. <https://i4replication.org/definitions.html>
- Ioannidis, J. P. A. (2012). Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), 645–654. <https://doi.org/10.1177/1745691612464056>
- Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, 127(605), F236–F265. <https://doi.org/10.1111/eoj.12461>
- Leamer, E. E. (1983). Let's Take the Con Out of Econometrics. *American Economic Review*, 73(1), 31–43.
- Long, J., & Ferrie, J. (2013). Intergenerational Occupational Mobility in Great Britain and the United States Since 1850. *American Economic Review*, 103(4), 1109–1137. <https://doi.org/10.1257/aer.103.4.1109>
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & Delgado López-Cózar, E. (2018). Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories. *Journal of Informetrics*, 12(4), 1160–1177. <https://doi.org/10.1016/j.joi.2018.09.002>
- Mattauch, L., Matthews, H. D., Millar, R., Rezai, A., Solomon, S., & Venmans, F. (2020). Steering the Climate System: Using Inertia to Lower the Cost of Policy: Comment. *American Economic Review*, 110(4), 1231–1237. <https://doi.org/10.1257/aer.20190089>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is Psychology Suffering from a Replication Crisis? What Does “Failure to Replicate” Really Mean? *American Psychologist*, 70(6), 487–498. <https://doi.org/10.1037/a0039400>
- Merton, R. K. (1973). The Normative Structure of Science. In N. W. Storer (Ed.), *The Sociology of Science*. The University of Chicago Press.
- Miao, B., & Zhong, S. (2015). Comment on “Risk Preferences Are Not Time Preferences”: Separating Risk and Time Preference. *American Economic Review*, 105(7), 2272–2286. <https://doi.org/10.1257/aer.20131183>
- Miguel, E. (2021). Evidence on Research Transparency in Economics. *Journal of Economic Perspectives*, 35(3), 193–214. <https://doi.org/10.1257/jep.35.3.193>
- Mirowski, P., & Sklivas, S. (1991). Why Econometricians Don't Replicate (Although They Do Reproduce). *Review of Political Economy*, 3(2), 146–163. <https://doi.org/10.1080/09538259100000040>
- Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Ozier, O. (2021). Replication Redux: The Reproducibility Crisis and the Case of Deworming. *The World Bank Research Observer*, 36(1), 101–130. <https://doi.org/10.1093/wbro/lkaa005>

- Peters, J., Langbein, J., & Roberts, G. (2018). Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity. *The World Bank Research Observer*, 33(1), 34–64. <https://doi.org/10.1093/wbro/lkx005>
- Peterson, D., & Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science*, 51(4), 583–605. <https://doi.org/10.1177/03063127211005551>
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. Taylor & Francis. <http://nukweb.nuk.uni-lj.si/login?url=http://search.ebscohost.com/login.aspx?authtype=ip&direct=true&db=nlebk&AN=143035&site=eds-live&scope=site&lang=sl>
- Roodman, D., & Morduch, J. (2014). The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence. *The Journal of Development Studies*, 50(4), 583–604. <https://doi.org/10.1080/00220388.2013.858122>
- Rothstein, J. (2017). Measuring the Impacts of Teachers: Comment. *American Economic Review*, 107(6), 1656–1684. <https://doi.org/10.1257/aer.20141440>
- Rubin, A., & Rubin, E. (2021). Systematic Bias in the Progress of Research. *Journal of Political Economy*, 000–000. <https://doi.org/10.1086/715021>
- Schafmeister, F. (2021). The Effect of Replications on Citation Patterns: Evidence From a Large-Scale Reproducibility Project. *Psychological Science*, 32(10), 1537–1548. <https://doi.org/10.1177/09567976211005767>
- Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable Publications are Cited More than Replicable Ones. *Science Advances*, 7(21), eabd1705. <https://doi.org/10.1126/sciadv.abd1705>
- Siler, K., Lee, K., & Bero, L. (2015). Measuring the Effectiveness of Scientific Gatekeeping. *Proceedings of the National Academy of Sciences*, 112(2), 360–365. <https://doi.org/10.1073/pnas.1418218112>
- Teplitskiy, M., Duede, E., Menietti, M., & Lakhani, K. R. (2022). How status of research papers affects the way they are read and cited. *Research Policy*, 51(4), 104484. <https://doi.org/10.1016/j.respol.2022.104484>
- Vazire, S., & Holcombe, A. O. (2021). Where are the Self-Correcting Mechanisms in Science? *Review of General Psychology*, 1–12. <https://doi.org/10.1177/10892680211033912>
- Vilhuber, L. (2019). Report by the AEA Data Editor. *AEA Papers and Proceedings*, 109, 718–729. <https://doi.org/10.1257/pandp.109.718>
- Vivalt, E. (2020). How Much Can We Generalize From Impact Evaluations? *Journal of the European Economic Association*, 18(6), 3045–3089. <https://doi.org/10.1093/jeea/jvaa019>
- von Hippel, P. T. (2022). Is Psychological Science Self-Correcting? Citations Before and After Successful and Failed Replications. *Perspectives on Psychological Science*, 174569162110725. <https://doi.org/10.1177/17456916211072525>
- Whaples, R. (2006). The Costs of Critical Commentary in Economics Journals. *Econ Journal Watch*, 3(2), 275–282.

- Whittington, L. A., Alm, J., & Peters, H. E. (1990). Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States. *American Economic Review*, 80(3), 545–556. <https://www.jstor.org/stable/2006683>
- Xie, Y., & Killewald, A. (2013). Intergenerational Occupational Mobility in Great Britain and the United States Since 1850: Comment. *American Economic Review*, 103(5), 2003–2020. <https://doi.org/10.1257/aer.103.5.2003>

The Online Appendix can be accessed on the OSF website under this link:
https://osf.io/kpywn/?view_only=2c4ecfdcc66f40459b6073ad9e7b42ce.

Appendix

Table A1: Citations on comments published in the AER between 2010 and 2020 and their respective original papers (OPs)

OP First Author / Comment First Author	OP				Comment		Citation Ratio (Comment/ OP) Since Comment Publication (in %)
	Since OP Publication		Since Comment Publication		Total Citations	Average Annual Citations	
	Total Citations (1)	Average Annual Citations (2)	Total Citations (3)	Average Annual Citations (4)			
Median	395	33	229	37	30	5	11.2
Mean	839	60	567	74	56	7	14.3
Acemoglu (2001) / Albouy (2012)	15,347	731	9,772	977	328	33	3.4
Alesina (2005) / Tella (2013)	1,912	112	1,277	142	15	2	1.2
Acemoglu (2008) / Cervellati (2014)	1,827	131	1,161	145	77	10	6.6
Koopman (2014) / Los (2016)	1,677	210	1,446	241	162	27	11.2
Deschênes (2007) / Fisher (2012)	1,652	110	1,423	142	389	39	27.3
Cerra (2008) / Mueller (2012)	1,503	107	1,195	120	23	2	1.9
Dahl (2012) / Lundstrom (2017)	1,409	141	802	160	12	2	1.5
Chetty (2014) / Rothstein (2017)	1,339	167	1,011	202	71	14	7.0
Ausubel (2004) / Okamoto (2018)	1,264	70	229	57	7	2	3.1
Plott (2005) / Isoni (2011)	1,129	66	813	74	218	20	26.8
Beaudry (2006) / Kurmann (2014)	1,070	67	656	82	40	5	6.1
Weyl (2010) / Tan (2018)	935	78	418	105	22	6	5.3
Mertens (2013) / Jentsch (2019)	850	94	491	164	42	14	8.6
Hatfield (2005) / Aygün (2013)	846	50	638	71	163	18	25.5
Fernández-V. (2011) / Born (2014)	826	75	703	88	48	6	6.8
Lustig (2007) / Burnside (2011)	769	51	652	59	186	17	28.5
Bohnet (2008) / Bolton (2010)	629	45	605	50	146	12	24.1
Abaluck (2011) / Ketcham (2016)	577	52	369	62	33	6	8.9
Andreoni (2012) / Cheung (2015)	519	52	424	61	60	9	14.2
Andreoni (2012) / Epper (2015)	519	52	424	61	44	6	10.4
Andreoni (2012) / Miao (2015)	519	52	424	61	66	9	15.6
Callen (2014) / Vieider (2018)	512	64	325	81	31	8	9.5
Davig (2007) / Farmer (2010)	502	33	440	37	58	5	13.2
Hastings (2004) / Taylor (2010)	483	27	373	31	56	5	15.0
Fehr (2001) / Petersen (2014)	470	22	170	21	27	3	15.9
Wright (2011) / Bauer (2014)	466	42	397	50	106	13	26.7
Muller (2009) / Fraas (2012)	417	32	371	37	21	2	5.7
Binsbergen (2012) / Schulz (2016)	413	41	313	52	41	7	13.1
Galí (2014) / Miao (2019)	395	49	168	56	18	6	10.7
Long (2013) / Hout (2013)	375	42	375	42	36	4	9.6
Long (2013) / Xie (2013)	375	42	375	42	100	11	26.7
Halevy (2008) / Saito (2011)	343	25	301	27	29	3	9.6
Adda (2006) / Abrevaya (2012)	296	19	207	21	23	2	11.1
Whittington (1990) / Crump (2011)	289	9	130	12	42	4	32.3
Feyrer (2017) / James (2020)	269	54	134	67	23	12	17.2
DeMarzo (2005) / Che (2010)	231	14	207	17	54	5	26.1
Brock (2013) / Krawczyk (2016)	229	25	181	30	16	3	8.8
Chang (2007) / Takahashi (2014)	203	14	124	16	13	2	10.5
Malmendier (2011) / Schneider (2016)	190	17	89	15	15	3	16.9
Jones (2014) / Caselli (2019)	187	23	103	34	35	12	34.0
Persson (2018) / Matsumoto (2018)	184	46	184	46	10	3	5.4
Weizsäcker (2010) / Ziegelmeyer (2013)	177	15	150	17	14	2	9.3
Steinsson (2008) / Iversen (2014)	171	12	104	13	13	2	12.5

Selten (2008) / Brunner (2011)	169	12	144	13	57	5	39.6
Mazzocco (2012) / Shrinivas (2018)	169	17	67	17	6	2	9.0
Kurmann (2013) / Cascaldi-Garcia (2017)	165	18	121	24	14	3	11.6
Demichelis (2008) / Heller (2014)	160	11	84	11	10	1	11.9
Bonjour (2003) / Amin (2011)	148	8	89	8	15	1	16.9
Echenique (2015) / Doğan (2017)	147	21	128	26	7	1	5.5
Coibion (2015) / Gagnon (2017)	137	20	108	22	14	3	13.0
Crainich (2013) / Ebert (2013)	125	14	125	14	32	4	25.6
Blonigen (2002) / Kelly (2010)	116	6	79	7	8	1	10.1
Armenter (2014) / Blum (2016)	110	14	75	13	3	1	4.0
Lemoine (2017) / Mattauch (2020)	56	11	28	14	10	5	35.7
Zhao (2008) / Chen (2010)	41	3	40	3	10	1	25.0
Fang (2017) / Matsumoto (2020)	40	8	28	14	2	1	7.1

Notes. Some comments are cited as discussion papers prior to their publication in the *AER*. In this table, we only include citations after their publication. However, only four out of the 56 comments have at least 10 citations prior to the *AER* publication, probably because discussion paper versions had circulated before: Albouy (2012), Burnside (2011), Cheung (2015), and Fisher et al. (2012). The same table can be found in Table D6 in the Online Appendix where we include the total citations of the reply as an additional column.