DEPARTMENT OF ECONOMICS AND FINANCE

SCHOOL OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

Do Replications Really Receive Fewer Citations?

Tom Coupé
Thomas Logchies
W. Robert Reed

WORKING PAPER

No. 15/2024

Department of Economics and Finance
UC Business School
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 15/2024

Do Replications Really Receive Fewer Citations?

Tom Coupé[1†]
Thomas Logchies[1]
W. Robert Reed[1]

October 2024

Abstract: This study examines the commonly held belief that replication studies suffer from a significant citation penalty compared to original research. Analysing a sample of 428 replications in economics, we challenge the traditional method of comparing replication citation rates directly with the originals they replicate. We introduce three incentive-based metrics to assess replication citations: the 'same original counterfactuals,' 'same author counterfactuals,' and 'same issue counterfactuals.' These metrics provide a more appropriate comparison by matching replications with comparable non-replication papers in terms of subject area, authorship, and publication venue. Our findings suggest that replications do not invariably receive fewer citations; rather, the citation gap narrows substantially or even reverses when using alternative counterfactual measures. This study not only reframes the discussion about the value of replication in economics but also encourages the academic community to reassess the incentives for conducting and publishing replication studies.

Keywords: Replications, Citations, Incentives, Academic Publishing

JEL Categories: A10, A14, B41, C80

[1] Department of Economics and Finance, University of Canterbury, NEW ZEALAND

† Corresponding author: Tom Coupé. Email: tom.coupe@canterbury.ac.nz

1. Introduction

There is substantial evidence of a 'replication crisis' in economics, with many published economics articles being unable to be replicated (Dewald, Thursby et al. (1986), Duvendack, Palmer-Jones et al. (2015), McCullough, McGeary et al. (2006), McCullough, McGeary et al. (2008), Camerer, Dreber et al. (2016), Chang and Li (2022)). While the number of papers that have been replicated is steadily growing, the share of papers that are replicated remains very small: Mueller-Langer et al. (2019) find that about 0.1% of publications published between 1974 and 2014 in the top 50 economics journals were replication studies.

Given replicability is a 'critical part of the scientific method' (Vilhuber 2020), many interventions to increase the rate of replications have been proposed. Some of these proposals -- such as journals requiring authors of published articles to submit their data and code (Galiani et al. (2017), McCullough (2009), Mueller-Langer et al. (2019), Zimmermann (2015)) -- focus on reducing the costs of replication. Other proposals focus on increasing the benefits of replications through, for example, having journals reserve space for them (Byington and Felps (2017), Tipu and Ryan (2021)) or making a replication a required part of PhD dissertations (Ryan and Tipu, 2022).

The benefits of producing or publishing replications are often assumed to be low because of an alleged replication citation penalty. Replications are believed to be cited less than "original" papers. This makes journal editors wary of publishing replication papers and authors reluctant to invest in doing them (Duvendack et al. (2015), Feigenbaum and Levy (1993), Koole and Lakens (2012), Byington and Felps (2017), Zimmermann (2015))[1].

[1] Throughout this paper, we assume that authors and editors are interested in citations and citations only. While this is obviously an abstraction of reality, citations are key in promotion considerations and journal rankings.

Many assume that replications receive fewer citations than original research. The traditional approach to calculating the replication citation penalty compares replications with the papers they replicate. Hubbard and Armstrong (1994) find that original marketing papers receive more citations than the extensions of those papers; while Ankel-Peters et al. (2023) report that comments in the *American Economic Review* receive fewer citations than the papers they criticize. In contrast, Marsden et al. (2018) find that citations to replication papers are higher than the average number of citations of articles published in linguistics and education journals.

The approach of this study is to investigate the relative citations of replications within a framework of incentives for authors to write -- and journals to publish -- replication papers. We develop three new measures that are incentive-based. We then compare the results from these measures with the traditional way of computing the replication citation penalty.

Using the traditional measure to compute the replication citation penalty, we confirm the findings of the prior literature with a sample 428 replications in economics. We find that replication studies receive only one-tenth of the citations of the original studies they replicate. However, our analysis argues that this traditional way of computing the citation penalty for replications is misleading. Alternative methods that better capture the counterfactual of a replication show a much smaller penalty. Some even find that replication papers receive more citations than non-replication papers.

Unfortunately, we do not have enough information about all the costs and benefits authors and journals face to compute the 'true' returns to writing and publishing replication versus original papers. Nevertheless, our findings challenge the commonly held belief that authors and journals face a substantial penalty for writing and publishing replications. At least in terms of citations, many replication papers garner the same or more citations as would be expected from other papers the authors and journals could have produced or published.

2. How should one compare citations of replications and original papers?

That replications studies get cited less is one of the "common beliefs about replication" (Nguyen, McKenzie et al. 2023). Table 1 gives examples from papers that make this claim.

Table 1
Papers that Claim a Replication Citation Penalty

Paper	***Quote***
Feigenbaum and Levy (1993, p. 218)	"Given the lower expected citation return to positive and negative replication studies versus original work, a citation maximizing researcher should produce replications only when they are of significantly lower cost (primarily in terms of time) to complete."
Koole and Lakens (2012, p611)	"Currently, replications receive few citations, because academics most often cite research that breaks new grounds. When replications are rarely cited, researchers have little to gain from conducting and publishing replication research."
Zimmerman (2015, p.3)	"Replications are less citable."
Duvendack et al (2015, p.170)	"Journal editors are concerned with per-page citations, and replications are thought to perform less well than original studies."
Byington and Felps (2017, p.154)	"As such, Hollenbeck advises authors that "straight replication, which is valuable but not urgent in the eyes of most editors… will not be viewed as generating new knowledge nor will it generate a large number of citations—which of course, is a problem for the editor" (2008: 23).""

The belief is so pervasive that some scholars have recommended a 'co-citation' requirement: journals should require that papers citing the original study also cite its replication (Coffman and Niederle (2015), Koole and Lakens (2012)).

Where this belief comes from is not exactly clear. One early piece of evidence, consistent with this belief, comes from Hubbard and Armstrong (1994). Analysing replications in the field of marketing, they noted *"of the 18 extensions in our sample that permitted an evaluation, 4 of*

the original studies were cited an average of 2.1 times per year, compared with the extensions' citation rate of 0.9. Moreover, the citation rates of the extensions were lower for 15 of the 18 comparisons and tied in two cases." More recently, Ankel-Peters et al. (2023), focusing on the *American Economic Review,* report *"that comments are cited on average seven times per year since their publication – compared to an average 74 citations per year for the original paper since publication of the comment."*

But are comparisons of citations of originals and replications the best way to identify a replication citation penalty? We next focus on how one would design an experiment if the goal was to obtain an unbiased measure of the replication citation penalty.

The ideal experiment would involve many authors. Each would write one replication and then spend the same amount of time to write one original paper. They would then submit the papers to journals. After publication, one would track citations to the two types of papers and compute whether there was a difference in their rates of citation.

The traditional way of computing the citation penalty does not mimic this ideal experiment. Typically, the authors of the replication paper are very different from the authors of the original study. Mueller-Langer et al. (2019) find replications tend to focus on high-impact research from renowned researchers and institutions. Therefore, the original papers used for the computation of the replication citation penalty are unlikely to be the appropriate counterfactuals for authors deciding to write a replication, or journals deciding to publish one. Accordingly, we propose three alternative counterfactuals.

Our first counterfactual records the citations received by all papers that cited the original study and were published in the same year that the replication was published. By focusing on papers that cited the original study, we attempt to match the subject area of the constructed counterfactual with that of the replication study. This is important because different subject

fields can have different citation rates. We call this approach the 'same original counterfactual'. In comparing the citations of replications and their 'same original counterfactuals", we note that the distributions of citations tend to be skewed. Accordingly, we report the median as well as the mean of the respective citation counts.

While the above approach tries to keep the subject area fixed, an alternative approach focuses on authors. It may be that the authors of 'same original counterfactuals' are more (or less) established scholars than the authors of replication studies. This would affect any comparison of citations across the two sets of authors. To address this issue, we next focus on non-replication papers written by the same authors who published the replication studies and published in the same year. We call this the 'same author counterfactual'.

Note that the counterfactual approaches above only capture the relative costs and benefits of replications in terms of citation rates. For example, when considering the incentives facing an author to produce a replication versus an original study, these approaches ignore the amount of work involved. If it is easier to produce a replication study, a researcher might find it preferable to do so even if they expected to receive fewer citations from the replication.

A second shortcoming of these approaches is that they only focus on papers that have already been published. They ignore differences in the probabilities of getting published that may exist between replication and original studies. There is ample evidence that not all working papers are published. For example, Baumann and Wohlrabe (2020) find evidence that a quarter of economics working papers are never published. A (non-random and multi-disciplinary) survey by the journal *Nature* (Baker (2016)) further found that, of those authors that tried to publish a replication, only about a half to two-thirds succeeded. Difference in publication rates is yet another reason why citation data might not capture the full set of incentives affecting researchers' decisions to produce replication studies.

So far, we have focused on counterfactuals that mainly pertain to authors. We next look at the decision to publish a study from the perspective of journals. Assuming journals only care about citations, our third counterfactual compares the citations of replications to the citations of other papers published in the same journal and issue as the replication paper. We call this the 'same issue counterfactual'[2].

3. Data and Analysis

The replications included in this study come from two websites that collect data on replications in economics: The Replication Network and ReplicationWiki (Höffler 2017). Since we are interested in analysing citations, we restrict the sample to replications published before 2022 so that we have at least 2.5 years of citations for each paper. Scopus identification numbers (EIDs) were obtained for each replication paper and each original paper. We were able to find Scopus EIDs for 428 original-replication pairs, and these form the basis of the analysis in this paper.

Our starting point is the traditional counterfactual, a simple comparison of the citations of replications and the original papers they replicated. Citations counts for both are taken starting from the year the replication was published. All citation counts are divided by the number of years since publication, so we compare annual citation rates.

[2] If a volume has several issues, we use the specific issue in which the replication was published. If there is only one volume and no issue, we use the volume.

Table 2: Comparing Citations of Replications and Originals, Using the Traditional Counterfactual

	(I) Replication citations	(II) Original citations	(III) Ratio
Mean	3.13	34.32	0.35
Median	1.37	13.33	0.11
Min	0.00	0.02	0.00
Q.1	0.50	5.78	0.04
Q.3	3.30	33.17	0.24
Max	78.05	650.50	65.00
# Obs.	428	428	428

Notes: this table is based on 428 pairs of replication papers and the papers they replicated. It compares citations of replications and citations of the papers they replicated, since the time the replication was published. Q1 and Q3 are the first and third quartile.

Table 2 compares the distributions of citations for replications and the originals they replicated. It shows that originals (column II) receive, on average, about 10 times more citations. The mean (median) of citations received by the replicated originals is about 34 per year (13 per year). The mean (median) of citations received by the replications (column I) is about 3 (1) citations per year.

Unlike the first and second columns, which treat replications and originals as independent, the third column ('Ratio') focuses on the match between each replication and the original it replicated. More specifically, it gives the distribution of the ratio of the replication paper's citations divided by the citation of the matched, original paper it replicated.[3]. It shows that on average, the replication citation count is about one third of the corresponding original's citation count (0.35). At the median, the ratio is only about 11%; that is, half of replication papers get more (less) than one-ninth of the citations of the originals they replicate (counting citations from the time of the publication of the replication). Interestingly, one replication actually received 65 times as many citations as the paper it replicated. Further analysis shows that 97.4% of the replications in our sample received fewer citations than the originals they replicated.

These findings are consistent with Hubbard and Armstrong (1994) and Ankel-Peters et al. (2023). However, this paper argues that this traditional way of computing the replication citation penalty does not identify the appropriate counterfactual. We therefore next explore our alternative measures of the replication citation penalty.

We first focus on the 'same original counterfactuals'; namely, those papers that cite the original paper in the same year as the replication. As discussed above, this approach attempts to match the replication paper with original papers in the same subject field(s).

[3] Column III is not the ratio of column I and II. Since the replication with average cites is not necessarily linked to the original with average cites, the average of the ratio is different from the ratio of the averages.

Table 3: Comparing Citations using the Same Original Counterfactual

	(I) Replication citations	(II) Same Original, Average Citations	(III) Same Original, Median Citations	(IV) Ratio, Same Original, Average Citations	(V) Ratio, Same Original, Median Citations
Mean	3.32	4.18	2.12	1.12	2.19
Median	1.50	3.27	1.67	0.50	0.94
Min	0.00	0.08	0.08	0.00	0.00
Q.1	0.71	2.11	1.08	0.20	0.39
Q.3	3.60	5.03	2.55	1.11	2.06
Max	78.05	37.28	23.87	25.00	35.31
# Obs.	395	395	395	395	395

Notes: this table is based on 395 pairs of replication papers and the papers they replicated. It compares citations of replications and citations of the papers that cited the paper they replicated, in the year the replication was published. Since each replication is matched to several counterfactuals, we explore both the distribution when using the mean counterfactual for each replication (column II and IV), and the median counterfactual for each replication (column III and V) The Ratio columns divide replication citations by the corresponding counterfactual citations. This sample excludes the 6 replications whose counterfactual had zero citations and the 27 replications who had original papers that were not cited by other papers in the year of the publication of the replication. Q1 and Q3 are the first and third quartile.

Column I of Table 3 essentially reproduces the analysis of column I of Table 2 except that the sample of replications consists of 395 replications rather than 428 replications. The difference arises because some of the matched originals were not cited by any other papers in the year of publication of the replication. Column II of Table 3 is different from its Table 2 analogue because each replication now has multiple counterfactuals. In Table 2, the "counterfactual" was the original study. In Table 3, the counterfactual for each replication consists of all the papers that cited the original study in the same year that the replication was published. Thus, each replication potentially has multiple counterfactuals.

Column II takes the average citation count for each set of counterfactuals and reports on the distribution of these averages. Using this sample of averages still produces a replication citation penalty, but much smaller than in Table 2. The mean citation count for the distribution of 395 averages of the 'same original counterfactuals' is 4.18 citations per year. This compares to an average citation count of 3.32 citations per year for the replications.

As mentioned before, citations are very skewed, so using the average of each set of 'same original counterfactuals' can be misleading. For this reason, column III reports on the distribution of medians of the 395 sets of 'same original counterfactuals'. We argue that medians provide a better indicator of the "typical" counterfactual for each replication.

This changes the picture quite dramatically: the average citation count for replications is almost 50% more than the average, median citation count for sets of 'same original counterfactuals'. (3.32 versus 2.12). In other words, if you compare the mean citation count for replications to the mean citation count of 'typical' counterfactuals (as measured by the median), replications are more likely to be cited. On the other hand, if you compare the median of the distribution of citations for replications, with the median of the distribution of "typical" counterfactuals (where "typical" is measured by the median of the matched set of counterfactuals), the

counterfactuals have a slight advantage (1.67 versus 1.50). Thus, using a more appropriate sample of counterfactuals either reverses the replication citation penalty or greatly reduces it, compared to using the citations of the original, replicated studies as the reference group.

Columns IV and V takes the comparison even further. Rather than comparing the distributions of citations and counterfactuals as independent distributions, it focuses on a direct comparison of replications with their matched sets of counterfactuals. Column IV compares the citations of each replication with the average citations from the corresponding matched set of counterfactuals. Column V compares the citations of each replication with the median citations from the corresponding matched set of counterfactuals. We then take the ratio of replication citations to the respective measure of counterfactual citations. A ratio larger than 1 for a given observation indicates that the replication receives more citations than its representative counterfactual. Columns IV and V report on the distribution of respective ratios.

Skipping over Column IV for the moment, Column V presents a generally favourable picture of the citations of replications compared to their matched counterfactuals. Using the median of the set of matched counterfactuals as a measure of the "typical" counterfactual for each replication, it shows that replications receive, on average, 2.19 times as many citations as their matched counterfactuals. Alternatively, if we use the median of the distribution of ratios (0.94) to compare replications, we find that replications and their matched counterfactuals have roughly equivalent citations. Relatedly, 48.4% of replications have the same or more citations than the median of their matched counterfactuals.

Column IV presents a similar though somewhat less favourable comparison of replications and their counterfactuals. On average, replications receive more citations (1.12) than their matched, "typical" counterfactuals, where "typical" is now defined as the average of the set of matched counterfactuals. However, at the median of the distribution of ratios, replications only receive

half as many citations (0.50) as their matched, "typical" counterfactuals. And only 28.1**%** of replications receive more citations than the average of their matched counterfactuals..

To summarize up to this point, the 'same original counterfactual" approach either reverses or greatly reduces the replication citation penalty compared to using the original, replicated article as a counterfactual. However, it is not without shortcomings. It fails to hold constant the citation productivity of authors. We therefore next compare citations of replication papers to citations of non-replication papers by the same authors that were published in the same year the replication was published.

Table 4 reports the results of this comparison. The replication studies comprising column I are somewhat different from previous tables. First, some replications were dropped because the author of the replication did not publish any non-replication papers in the same year. Further, replications with multiple co-authors are included more than once if the co-authors separately published multiple non-replication papers in the same year. This yielded 495 replications for comparison.

Column II of Table 4 is similar to column II of Table 3, except that the set of counterfactuals for any given replication consists of non-replication papers by the same author published in the same year as the replication. Column III of Table 4 uses the same set of counterfactuals as column II, but selects the median from the set of a replication's counterfactuals, rather than the average. As most counterfactual sets consist of only one paper, using the average or the median provides similar measures of a typical, "same author counterfactual".

A comparison of column I of Table 4 with columns II and III shows that replication papers tend to get citation counts similar to the citation counts of non-replication papers by the same author. Using the mean of the respective distributions finds that replications get slightly more

citations than their ‘same author counterfactuals’ (3.90 versus 3.68 and 3.23). Using the median finds that replications get slightly fewer citations (1.55 versus 2.07 and 1.81).

Columns IV and V directly compare replication citations with citations of their matched set of ‘same author counterfactuals” by taking the ratio of the former over the average/median of the latter. The average of the distribution of ratios is 2.33 and 3.27, respectively, which indicates a citation advantage of replications over their non-replication counterfactuals since the ratios are appreciably larger than 1. The median of the distributions finds a slight replication citation penalty, with the median replication producing 0.79 and 0.95 citations to a single citation by the respective counterfactual. In conclusion, the ‘same author counterfactual’ approach, much like the ‘same original counterfactual’ approach, either reverses or substantially diminishes the replication citation penalty compared to using the original, replicated article as a counterfactual.

Table 5 collects our findings from previous tables and adds standard errors for the ratios of replication citations to counterfactual citations. A ratio of 1 means replications and their counterfactuals receive an equal number of citations, so our significance tests will focus on whether a ratio is significantly different from 1 (not the usual different from 0). Given that citation counts tend to have outliers with extreme values, we analyse median (the even columns) as well as mean ratios (the odd columns).

The standard errors come from regression analyses. For the mean, we get the standard error from an OLS regression of the ratio on a constant. For the median, we use quantile regression (for the median quantile). The estimated coefficient of the constant term reflects the mean (median) ratio and the corresponding standard error can be used to test whether that mean (median) is statistically different from 1.

Table 4: Comparing Citations using the Same Author Counterfactual

	(I) Replication citations	(II) Same Author, Average Citations	(III) Same Author, Median Citations	(IV) Ratio, Same Author, Average Citations	(V) Ratio, Same Author, Median Citations
Mean	3.90	3.68	3.23	2.33	3.27
Median	1.55	2.07	1.81	0.79	0.95
Min	0.00	0.03	0.03	0.00	0.00
Q.1	0.71	0.93	0.75	0.32	0.34
Q.3	4.00	4.00	3.65	2.20	2.63
Max	78.05	55.57	55.57	100.00	125.00
# Obs.	495	495	495	495	495

Notes: this table is based on 495 authors of replication papers. It compares citations of their replication paper and the citations of the non-replication papers they published in the year the replication was published. The Ratio columns divide replication citations by the corresponding counterfactual citations. This sample excludes the 13 authors whose counterfactual had zero citations. Authors of replications who had no non-replication papers that were published in the year of the publication of the replication are excluded. Since each replication is matched to several counterfactuals, we explore both the distribution when using the mean counterfactual for each replication (column II and IV), and the median counterfactual for each replication (column III and V). Q1 and Q3 are the first and third quartile.

Table 5: Mean/Median of Ratio of Replication Citations to Counterfactual Citations - Counterfactuals relevant to Authors

	Traditional Counterfactual			Same Original Counterfactual			Same Author Counterfactual			
	(I) Mean, Citation Ratio	(II) Median, Citation Ratio	(III) Mean, Mean Citation Ratio	(IV) Median, Mean Citation Ratio	(V) Mean, Median Citation Ratio	(VI) Median, Median Citation Ratio	(VII) Mean, Mean Citation Ratio	(VIII) Median, Mean Citation Ratio	(IX) Mean, Median Citation Ratio	(X) Median, Median Citation Ratio
Mean/Median	0.35	0.11	1.12	0.50	2.19	0.94	2.33	0.79	3.27	0.95
Standard Error	0.15	0.01	0.12	0.04	0.21	0.07	0.27	0.06	0.44	0.07
T-Stat	-4.30	-102.57	1.01	-13.81	5.77	-0.80	4.88	-3.84	5.21	-0.68
P-value	0.00	0.00	0.31	0.00	0.00	0.43	0.00	0.00	0.00	0.49

Notes: this table provides means and medians of the ratio of replication citations to counterfactual citations, for various counterfactuals relevant to authors. It also provides the standard errors of these means and medians which allows to analyze whether the ratio is statistically significantly different from one. The T-Stat and P-value rows give the t-statistic and p-value of this test. Columns I and II are based on table 2, columns III to VI on table 3, and columns VIII to X on table 4.

Table 5, columns I and II, shows that the citation penalty is large and statistically significant when one uses the traditional approach to comparing a replication with the original paper it replicated. In contrast, the picture changes markedly when using either of the two alternative counterfactuals (columns III to X).

If we take the most preferred counterfactuals to be comparisons that use the median from the set of matched counterfactuals, there are four columns to focus on: Columns V and VI ("Same Original Counterfactual") and columns IX and X ("Same Author Counterfactual"). In columns V and IX, the mean ratios of replication citations to their matched counterfactuals is 2.19 and 3.27. Both indicate a substantial citation benefit to publishing a replication compared to their respective counterfactuals. Both mean ratios are significantly different from one (where 1 represents equality in citations for replications and their counterfactuals). In columns VI and X, the median ratios of replication citations to their matched counterfactuals are 0.94 and 0.95. These indicate a slight replication citation penalty but the median ratios are not statistically different from one.

Columns III and IV ("Same Original Counterfactual") and columns VII and VIII ("Same Author Counterfactual") use the means of the sets of matched counterfactuals to represent the "typical" alternative to a replication study. Here the picture is murkier. Columns III and VIII both indicate a citation benefit for replications, with mean ratios of 1.12 and 2.33. But only the latter ratio is significantly different from one. Citation benefits become citation penalties in columns IV and VIII (0.50 and 0.79) when the comparison is based on the median ratio comparing replications with a "typical" (mean) counterfactual. However, only the latter ratio is significantly different from one.

In summary, while the traditional approach indicates a substantial citation penalty for replications, we find mixed evidence using measures that more closely align with the incentives

facing authors. Three of the five statistically significant results find a replication citation benefit rather than a penalty (columns V, VII, and IX).

So far we have focused on estimating the citation penalties for authors. We now study the incentives for journals to publish replications. To do that, we compare the citations of replications to the citations of other papers published in the same issue in which the replication was published. In particular, we address the question, Is there a citation penalty for journals publishing replications rather than non-replication papers?

Column I of Table 6 reports the distribution of citations for replications, reproduced from column I of Table 2. Columns II and III of Table 6 compare that distribution with the distributions for the respective same issue counterfactuals, respectively using the mean of the set of matched counterfactuals and the median of the set of matched counterfactuals to identify the "typical" counterfactual.

If we use the median of the set of matched counterfactuals to represent the "typical" same issue counterfactual (cf. Column III), we scarcely see a difference. Mean annual citations for replications is 3.13, while mean annual citations for the typical, same issue counterfactual is 3.25. If we focus on the median of the two distributions, we see a slightly larger penalty for replications (1.37 versus 1.88). If instead we use the average of the set of same issue counterfactuals to represent the typical counterfactual, the penalty becomes larger: 3.13 versus 5.09 (Mean) and 1.37 versus 3.02 (Median).

Turning attention to the matched-pair ratios of columns IV and V, we find a slight citation benefit for replications (1.01 and 1.66). However, as we have seen in previous tables, when we compare the medians of the distributions of matched-pair citation ratios, replication papers get about half to 3/4th of the citations of their counterfactuals. About 60% (using the median to identify the typical counterfactual, column V) and 74% (using the mean to identify the typical

counterfactual, column IV) of replications get cited less than their matched, typical counterfactuals.

Journals, however, not only control whether they publish a replication paper, they also determine how much journal space (i.e. pages) to allocate to a paper. Hence, they might be interested in the number of citations per page rather than the number of citations per article. Table 7 shows the distribution of per page citations for replications and papers published in the same issue as the replications. Given replications tend to be shorter, the per page statistics are more favourable to replications. Mean citations per page for replications is greater than mean citations per page for either of the two measures of a "typical", same issue counterfactual. The other per page measures show either a larger replication citation benefit or a smaller replication citation penalty compared to the total citation counts of Table 6.

Table 6: Comparing Citations using the Same Issue Counterfactual

	(I) Replication citations	(II) Same Issue, Average Citations	(III) Same Issue, Median Citations	(IV) Ratio, Same Issue, Average Citations	(V) Ratio, Same Issue, Median Citations
Mean	3.13	5.09	3.25	1.01	1.66
Median	1.37	3.02	1.88	0.50	0.75
Min	0.00	0.09	0.01	0.00	0.00
Q.1	0.50	1.44	0.82	0.18	0.29
Q.3	3.30	7.11	4.44	1.03	1.66
Max	78.05	35.25	23.27	29.12	47.08
# Obs.	428	428	428	428	428

Notes: this table is based on 428 papers. It compares citations of their replication paper and the citations of the non-replication papers published in the same issue the replication was published. The Ratio columns divide replication citations by the corresponding counterfactual citations. Since each replication is matched to several counterfactuals, we explore both the distribution when using the mean counterfactual for each replication (column II and IV), and the median counterfactual for each replication (column III and V). Q1 and Q3 are the first and third quartile.

Table 7: Comparing Citations per Page using the Same Issue Counterfactual

	(I) Replication Citations per Page	(II) Same Issue, Average Citations per Page	(III) Same Issue, Median Citations per Page	(IV) Ratio, Same Issue, Average Citations per Page	(V) Ratio, Same Issue, Median Citations per Page
Mean	0.26	0.24	0.15	1.39	2.30
Median	0.12	0.16	0.11	0.77	1.12
Min	0.00	0.01	0.00	0.00	0.00
Q.1	0.05	0.08	0.05	0.32	0.53
Q.3	0.27	0.31	0.20	1.67	2.46
Max	3.90	2.15	1.04	35.52	53.27
# Obs.	417	417	417	417	417

Notes: this table is based on 417 replication papers. It compares citations per page of their replication paper and the citations per page of the non-replication papers published in the same issue the replication was published. The Ratio columns divide replication citations by the corresponding counterfactual citations. Since each replication has multiple counterfactuals we show results with both the mean and median of these counterfactuals. For some papers, Scopus had no information about the number of pages. Since each replication is matched to several counterfactuals, we explore both the distribution when using the mean counterfactual for each replication (column II and IV), and the median counterfactual for each replication (column III and V). Q1 and Q3 are the first and third quartile.

Table 8: Mean/Median of Ratio of Replication Citations to Same Issue Counterfactual Citations - Counterfactuals Relevant to Editors

	Traditional Counterfactual			Same Issue Counterfactual			Same Issue Counterfactual, per Page			
	(I) Mean, Citation Ratio	(II) Median, Citation Ratio	(III) Mean, Mean Citation Ratio	(IV) Median, Mean Citation Ratio	(V) Mean, Median Citation Ratio	(VI) Median, Median Citation Ratio	(VII) Mean, Mean Citation Ratio	(VIII) Median, Mean Citation Ratio	(IX) Mean, Median Citation Ratio	(X) Median, Median Citation Ratio
Mean/Median	0.35	0.11	1.01	0.50	1.66	0.75	1.39	0.77	2.30	1.12
Standard Error	0.15	0.01	0.10	0.04	0.17	0.05	0.12	0.05	0.21	0.08
T-Stat	-4.30	-102.57	0.14	-13.93	3.83	-4.82	3.36	-4.48	6.27	1.61
P-value	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.11

Notes: this table provides means and medians of the Ratio of Replication Citations to Counterfactual Citations, for various counterfactuals relevant to editors. It also provides the standard errors of these means and medians which allows to analyze whether the ratio is statistically significantly different from one. The T-Stat and P-value rows give the t-statistic and p-value of this test. Columns I and II are based on table 2, columns III to VI on table 3, and columns VIII to X on table 4.

Table 8 repeats the analysis of Table 5, except it focuses on citations from the perspective of journals, using citations from non-replication studies that were published in the same journal issue as the replication paper. Columns I and II of Table 8 reproduce the mean and median ratios of replications to the original, replicated papers in Table 2 (cf. column III). Columns III to VI ("Same Issue Counterfactual") and columns VII to X ("Same Issue Counterfactual, per Page") report ratios using alternative, counterfactual measures that incorporate the publication incentives of the journals. The conclusions are similar to Table 5.

If we take the median of the set of matched same issue counterfactuals as the preferred opportunity cost of publishing a replication, and per page citations as being the citation indicator of most relevance for journals (cf. columns IX and X), then from the journals' perspective, there is a citation benefit, rather than a penalty, to publishing a replication. The mean (median) ratio of per page citations for replications compared to their respective "median" counterfactuals is 2.30 (1.12). While both values are greater than one, indicating a replication citation benefit, only the first value is statistically significant.

Of the remaining incentive-based counterfactual measures, five of the ratios are statistically significant (cf. columns IV to VIII). Two find a replication citation benefit (columns V and VII), and three find a replication citation penalty (columns IV, VI, and VIII).

Up to this point we have ignored how the findings of a replication might affect their citations, and thus their attractiveness to journals. While researchers who replicate a paper do not know in advance whether their replication will confirm the original results, journals are able to see the outcome of a replication. If a journal thinks that replications that fail to confirm the original study are more likely to be cited, this could bias journals towards publishing negative replications.

To investigate this issue, a team of research assistants led by one of the co-authors of this study classified each of the 428 replications in this study into three categories: confirmed, disconfirmed, and mixed/unclear. The team generally took the replicating author's own assessment of the outcome of their replication. For example, if a replicating author stated in the abstract or conclusion that their paper failed to confirm the original study, we did not second-guess their assessment. Each paper was independently reviewed multiple times by the research assistants, often two or more times. The co-author leading the team also independently assessed each study.

There was general agreement among the team when it came to assessing replications that confirmed the original study. Distinguishing between Mixed/Unclear and Disconfirmed was more challenging. For example, if a study investigated the role of an intervention on the energy consumption of poor households in Brazil, and a study identifying itself as a replication extended the study to poor households in Mexico, but obtained different results, did that disconfirm the original study? Or was it unclear how it should affect one's interpretation of the original study? This depended on the target population of the original study. Was the target population of the original study poor households in Brazil? Or poor households in a developing country in South America? Often the authors of the original studies did not identify their intended target populations.

Table 9 reports annual citations per article for each of the three categories of replications. Replications that confirm the original study have an average (median) citation rate of 1.99 (1.00) citations per year. Replications that disconfirm the original study, or report mixed/unclear results, are cited approximately 50% more. For Mixed/Unclear replications, the mean and median citations per year are 3.29 and 1.55. The citation rates for Disconfirmed replications are very similar, with mean and median citations per year equal to 3.50 and 1.50. As we have previously seen, however, unadjusted citation rates can be misleading. As a result,

we match the three sets of replications with their respective same issue counterfactuals and calculate per page citation rates.

Table 10 reports the associated results, using the median of the set of matched non-replication studies to represent a "typical" counterfactual. It is like columns IX and X of Table 8, except that, rather than reporting these for all replications, it does so separately for Confirmed, Mixed/Unclear, and Disconfirmed replications; Like columns IX and X of Table 8 it uses the metric of citations per page rather than total article replications.

The first thing to notice from Table 10 is that the relative citation advantage of Mixed/Unclear and Disconfirmed replications that was apparent in Table 9 disappears when citations are measured on a per page basis, and replications are matched against a "typical" counterfactual from the same journal issue. If anything, Confirmed replications do better relative to their typical counterfactuals than Mixed/Unclear and Disconfirmed replications. The mean and median ratios for Confirmed replications are 2.75 and 1.51. The corresponding values for Mixed/Unclear and Disconfirmed are 2.42 and 1.24, and 2.08 and 1.00. For journals interested solely in citations, these results indicate that they need not make publication decisions based on the outcome of submitted replications.

A related finding from Table 10 is that replications fare no worse, and generally better than their same issue, non-replication matches for all three categories of replication outcomes. In three of the six tests from Table 10, there is a replication citation benefit and it is statistically significant (cf. columns I, III, and V). In one case it is marginally significant (cf. column II: p-value = 0.06). The other two tests are statistically insignificant. For none of the replication outcomes do we estimate a replication citation penalty.

Table 9: Comparing Citations of Replications with Various Outcomes

	Replication citations - Confirmed	Replication citations - Mixed/Unclear	Replication citations - Disconfirmed
Mean	1.99	3.29	3.50
Median	1.00	1.55	1.50
Min	0.00	0.00	0.00
Q.1	0.33	0.67	0.64
Q.3	1.80	3.82	3.67
Max	24.17	22.90	78.05
# Obs.	89	109	230

Notes: this table is based on 428 replication. It compares replications on the basis of the outcome of the replication, whether the replication confirmed or disconfirmed the original paper, or had found mixed results. Q1 and Q3 are the first and third quartile.

Table 10: Comparing Citations of Replication with Various Outcomes, Using the Median Same Issue Counterfactual.

	Confirmed Replications		Mixed/Unclear Replications		Disconfirmed Replications	
	(I) Mean Ratio per Page	(II) Median Ratio per Page	(III) Mean Ratio per Page	(IV) Median Ratio per Page	(V) Mean Ratio per Page	(VI) Median Ratio per Page
Mean/Median	2.75	1.51	2.42	1.24	2.08	1.00
Standard Error	0.55	0.27	0.34	0.21	0.28	0.09
T-Stat	3.15	1.92	4.21	1.12	3.86	0.00
P-value	0.00	0.06	0.00	0.27	0.00	1.00

Notes: this table provides means and medians of the ratio of replication citations to counterfactual Citations, per page (like table 7). It also provides the standard errors of these means and medians which allows to analyze whether the ratio is statistically significantly different from one. The T-Stat and P-value rows give the t-statistic and p-value of this test. For some papers, Scopus had no information about the number of pages. Since each replication is matched to several counterfactuals, we explore both the distribution when using the mean counterfactual for each replication (column I, III and V), and the median counterfactual for each replication (column II, IV and VI). Column I and II use replications that confirm the original.Column III and IV restrict the sample to replications that have mixed/unclear results. Column V and VI restrict the sample to replications that disconfirm the original.

4. Conclusion

Do replications receive fewer citations than so-called "original" research? There is a common belief that they do. This paper makes two main points. First, the common belief rests on an inappropriate comparison. It stems from the huge difference in the number of citations between replications and the originals they replicated. Using a sample of 428 replications in economics published between 1958 and 2021, we too find that that original studies receive many more citations than the studies that replicated them: approximately ten times as many. However, we argue that these original studies are not the appropriate counterfactual for assessing whether replications receive fewer citations than other papers.

We develop three alternative approaches that better match the opportunity costs of producing a replication: two from the perspective of researchers ("same original counterfactuals" and ("same author counterfactuals") and one from the perspective of journals ("same issue counterfactuals"). The "same original counterfactuals" approach compares the citations of replications against the citations received by other, non-replication papers that cite the replicated original and are published in the same year as the replication. The "same author counterfactuals" approach compares citations of replications with citations of non-replications published by the same authors in the same year the replication was published. The "same issue counterfactuals" approach compares citations of replications with those from other, non-replication papers published in the same journal issue as the replication.

Based on these three approaches, we present a range of results. Our main results are summarized in Tables 5 and 8, with our preferred results measuring the opportunity cost of a replication as the median number of citations received by the set of matched counterfactuals. Using this measure, we find that the mean and median ratios of replication citations to counterfactual citations range from 0.75 to 3.27, with ratios greater (less) than 1 indicating a

replication citation benefit (penalty). Over all estimates, our largest citation "penalty" has replications being cited half as much as their counterfactuals. At the other end, our largest citation "benefit" has replications being cited over three times as much. We also investigated whether the outcome of a replication affected its citations when compared to non-replications published in the same journal issue. We found little difference in the citation ratio of replications to their counterfactuals based on outcome (confirmed, mixed/unclear, disconfirmed).

A limitation of our study is that we focus on citations. There are many dimensions to the decision to produce and publish a replication. Citations are but one dimension. Nevertheless, the existence of a "citation penalty" is a commonly mentioned reason for why we don't see more replications being produced, and more replications being published. It is hoped that this study will alleviate the concern by researchers and journals that there is a substantial citation penalty associated with the decision to produce and publish a replication.

References

Ankel-Peters, J., et al. (2023). "Is economics self-correcting? Replications in the American Economic Review."

Baker, M. (2016). "1,500 scientists lift the lid on reproducibility." Nature **533**(7604).

Baumann, A. and K. Wohlrabe (2020). "Where have all the working papers gone? Evidence from four major economics working paper series." Scientometrics **124**(3): 2433-2441.

Byington, E. K. and W. Felps (2017). "Solutions to the credibility crisis in management science." Academy of Management Learning & Education **16**(1): 142-162.

Camerer, C. F., et al. (2016). "Evaluating replicability of laboratory experiments in economics." Science **351**(6280): 1433-1436.

Chang, A. C. and P. Li (2022). "Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say â€ œOften Notâ€." Critical Finance Review **11**(1): 185-206.

Coffman, L. C. and M. Niederle (2015). "Pre-analysis Plans Have Limited Upside, Especially Where Replications Are Feasible." Journal of Economic Perspectives **29**(3): 81-98.

Dewald, W. G., et al. (1986). "Replication in empirical economics: The journal of money, credit and banking project." The American Economic Review: 587-603.

Duvendack, M., et al. (2015). "Replications in economics: A progress report." Econ Journal Watch **12**(2): 164-191.

Feigenbaum, S. and D. M. Levy (1993). "The market for (ir)reproducible econometrics." Social Epistemology **7**(3): 215-232.

Galiani, S., et al. (2017). Incentives for replication in economics, National Bureau of Economic Research.

Höffler, J. H. (2017). "ReplicationWiki: Improving transparency in social sciences research." D-Lib Magazine **23**(3): 1.

Hubbard, R. and J. S. Armstrong (1994). "Replications and extensions in marketing: Rarely published but quite contrary." International Journal of Research in Marketing **11**(3): 233-248.

Koole, S. L. and D. Lakens (2012). "Rewarding Replications:A Sure and Simple Way to Improve Psychological Science." Perspectives on Psychological Science **7**(6): 608-614.

Marsden, E., et al. (2018). "Replication in Second Language Research: Narrative and Systematic Reviews and Recommendations for the Field." Language Learning **68**(2): 321-391.

McCullough, B. D. (2009). "Open Access Economics Journals and the Market for Reproducible Economic Research." Economic analysis & policy **39**(1).

McCullough, B. D., et al. (2006). "Lessons from the JMCB Archive." Journal of Money, Credit and Banking: 1093-1107.

McCullough, B. D., et al. (2008). "Do economics journal archives promote replicable research?" Canadian Journal of Economics/Revue canadienne d'économique **41**(4): 1406-1420.

Mueller-Langer, F., et al. (2019). "Replication studies in economics—How many and which papers are chosen for replication, and why?" Research Policy **48**(1): 62-83.

Nguyen, P.-Y., et al. (2023). "Systematic reviewers' perspectives on replication of systematic reviews: A survey." Cochrane Evidence Synthesis and Methods **1**(2): e12009.

Ryan, J. C. and S. A. Tipu (2022). "Business and management research: Low instances of replication studies and a lack of author independence in replications." Research Policy **51**(1): 104408.

Tipu, S. A. A. and J. C. Ryan (2021). "Are business and management journals anti-replication? An analysis of editorial policies." Management Research Review **45**(1): 101-117.

Vilhuber, L. (2020). "Reproducibility and replicability in economics." Harvard Data Science Review **2**(4): 1-39.

Zimmermann, C. (2015). "On the need for a replication journal." FRB St. Louis Working Paper(2015-16).