

# Assessing the Rate of Replication in Economics

By JAMES BERRY, LUCAS C. COFFMAN, DOUGLAS HANLEY,  
RANIA GIHLEB AND ALISTAIR J. WILSON\*

Replications are a key component in the scientific process; helping the profession sift robust empirical findings from mistakes. However, while replications are desirable, there remains uncertainty over just how often they occur. Our paper is a coarse attempt to shed some light on this uncertainty. Focusing on all empirical papers in the *American Economic Review's* (AER) centenary volume, our main finding is that 29 percent of the papers were replicated (to some degree), while 59 percent had either been replicated or extended.

Our measurements here are complementary to two other papers in this issue: Sunkhtankar (2017), which examines replications of development papers; and Hamermesh (2017) which examines ten high-profile papers in labor. Where each of these papers examines a particular sub-field, our own work surveys all empirical works in a year's volume of a top general-interest journal

The definition of a "replication" is admittedly somewhat mercurial. In some fields a replication is an attempt to verify the original paper's results with the same data. For example, a graduate student might redo the analysis in order to better learn a technique and detect an error in the original code. Alternatively, a replication might reproduce the original paper's experiment in the field. For example, a treatment from the original paper could be used as a control in a follow-up paper that focuses on extending the original. To capture the variety of replication attempts, in our main coding exercise we take a top-level approach, defining a repli-

cation as any project that reports results that speak directly to the veracity of the original paper.<sup>1</sup>

Our first data exercise examines all published papers in top-200 economics journals that cite one of the 70 empirical papers in the 2010 *AER*. Our measurements of replication were manually coded for the 1,546 papers in our citing sample. Cognizant that our first exercise with its focus on published papers may be too narrow, our second exercise surveys the authors of the original papers and a subsample of the citing papers to measure their beliefs and awareness of replications in the larger literature.

## Manual Coding Sample

Our base measurement is for the *AER's* 100th volume, published in 2010. In total the volume contains 223 papers, where we then excluded two Nobel addresses and 119 papers from *Papers & Proceedings* to focus on peer-reviewed work. We then exclude a further nine articles that were comments or replies. This left us with 95 peer-reviewed articles across an array of fields. We further removed 25 papers that were purely theoretical in nature, as our focus was on empirical replications.

The final *Volume Sample* for which we measured the rate of replication was given by 70 empirical papers. We collated all published works citing a paper in our *Volume Sample* via *Web of Science* (WoS) in June of 2016. Every paper in the *Volume Sample* therefore had at least five years since publication to accrue citations. In total there were 2,945 citing papers. Restricting the citing papers to come from a top-200 economics journal (using WoS impact factors) lead to a final sample of 1,558

\* Berry: Cornell University. Coffman: Harvard University. Hanley: University of Pittsburgh. Gihleb: University of Pittsburgh. Wilson: University of Pittsburgh. We are thankful to Muriel Niederle for her advice and feedback on this project. In addition, we are grateful to a number of authors from our volume and citing samples that provided feedback and information on their works.

<sup>1</sup>As we were aware this approach leads to some subjectivity, we also measure replications using the narrower definitions in Clemens (forthcoming).

Table 1—: Citation counts for our Volume sample ( $N = 70$ )

	Mean	Min	Median	Max
<i>Google Scholar</i> (GS)	227.6	7	139	1,246
<i>Web of Science</i> (WoS)	42.1	1	28.5	195
<i>Top-200 Economics journal</i> (WoS-200)	22.3	0	15	108

citations, which we refer to as our *Citing Sample*.<sup>2</sup>

The published citation counts for *Volume Sample* papers had substantial variation. Table 1 provides summary statistics for the number of Google Scholar (GS) cites, the number of WoS citations, as well the top-200 Economics citations that we used to generate our *Citing Sample*. The median paper has 139 GS citations, 28.5 WoS citations, of which 15 are in top-200 economics journals.

The *Citing Sample* papers were divided among the projects' coauthors for coding, mostly along the lines of field so that for the large majority of cases the coders were specialists in the field. After accounting for a small number of citing papers which were not available (or in one case, not in English), a total of 1,546 papers were coded by the five coauthors.<sup>3</sup>

Every paper was coded according to: (i) the coder's subjective opinion on whether or not the paper was a replication of the relevant volume paper; (ii) the coders opinion on whether the paper was an extension of the relevant volume paper; and (iii) three variables reflecting whether the paper used the same statistical model/specification, used the same data sample, and/or used data drawn from the same population as the relevant volume paper. These final three variables were recorded so that we could encode the more-concrete definitions of a replication/robustness tests in Clemens (forthcoming).

<sup>2</sup>Each *Citing Sample* entry is more properly thought of as a directed edge between a citing paper and a *Volume Paper*, as some papers cite multiple *Volume Sample* works.

<sup>3</sup>One paper in the ? and one paper in the *Citing Sample* had authors in common with the present paper; neither were coded by the overlapping coauthor.

## Manual Coding Results

Of the 1,546 citing papers, 52 were coded as replications. So roughly three and a half out of every hundred citations contain content that replicates the original. Across the 70 *Volume Sample* papers, 29 percent (20 papers) have at least one citing paper coded as a replication. Conditional on being replicated, the average number of replications per paper is 2.6. Though most papers with replication attempts have very few—eleven have just one, and three have two—five papers (7 percent) have five or more replications.<sup>4</sup>

In addition to our replication coding, we also report results in Table 2 for three alternative measurements (both for the *Volume* and *Citing Samples*): (i) a *robustness* test à la Clemens (forthcoming);<sup>5</sup> (ii) an extension (subjectively coded by a coauthor); and (iii) *Any* of replication/robustness/extension.

In total 42 of the 70 volume papers have one or more citation coded as a replication/robustness/extension. Though this represents 60 percent of the empirical papers in the AER volume, the majority of this follow-up work is coming through robustness tests and extensions. Moreover, of the papers coded as being replications,

<sup>4</sup>Replication work is typically independent, not being produced by the original authors. Forty-eight of the 52 replicating papers (92 percent) have no authors in common with the original volume paper.

<sup>5</sup>Across the 1,546 coded citing papers we find: no verifications (using the same data and econometric specification); two reproductions (the same econometric specification but a with a new dataset drawn from the same population); 57 reanalyses (altered econometric specifications on the same data or population); and 16 extensions (same specification on a different population). Robustness tests are defined as Clemens' reanalyses or extensions.

Table 2—: Coding Rates

	Replications	Robustness	Extension	Any
<i>Volume Sample</i>	28.6%	40.0%	48.6%	60.0%
	(20)	(28)	(34)	(42)
<i>Citing Sample</i>	3.4%	4.7%	7.8%	11.0%
	(52)	(73)	(121)	(170)

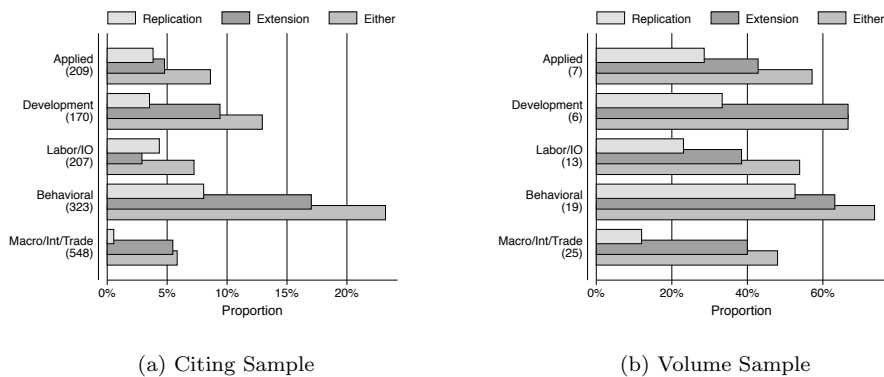


Figure 1. : Replications/Extensions by Field

none are pure replications, papers explicitly constructed to examine the veracity of the original paper’s results.

Figure 1(a) breaks out the rate of replication and extension (at least one) by volume paper field. Though sample sizes are small, some patterns emerge. Just over half of the Behavioral/Experimental papers in the volume had at least one replication attempt. All other fields saw a published replication attempt on between 12 and 33 percent of the *Volume Sample*. In contrast to replications, citing papers extend the original work across all fields at much higher rates, with between 38 and 67 percent of the volume papers extended. Figure 1(b) provides a parallel illustration for the *Citing Sample*, indicating the fraction of citing papers (by field) that are replications or extensions.

Volume paper replications accrue uniformly across the measured period. This is illustrated in Figure 2(a), which shows the cumulative fraction of volume papers with one or more replication from 2010 (the volume’s publication year) to 2016 (the year we collected data on citations).

One consistent predictive variable for whether a particular volume paper is replication is the number of times that paper is cited. In Figure 2(b) we illustrate the effect by graphing the CDFs for WoS citations for papers with no replications and those with one or more. Here the figure clearly illustrates the stochastic ordering of the data.<sup>6</sup> In particular, the figure shows that all papers in our sample with more than 100 published citations have replications. That is, for important results, the profession does a better job at replicating findings.

### Survey Sample

The previous analysis measures replication attempts through the subjective judgment of this paper’s five coauthors, researchers in the field reading published work in Economics. Though a starting point, the estimates may err for a couple

<sup>6</sup>The marginal effect from a probit estimation suggests that ten additional published citations increases the chance that a volume paper has a replication by 5.2 percent (significant at the 1 percent level).

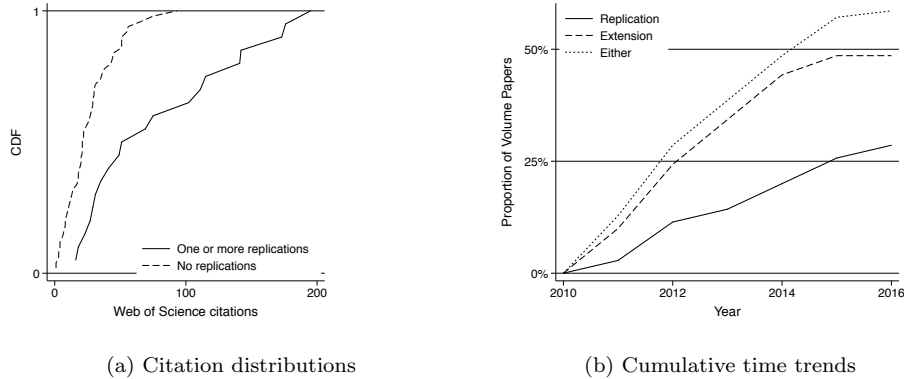


Figure 2. : Volume paper replications: Time and Citations

of reasons. First, there may be replications not in the sample—unpublished work, papers not in Economics, undistributed graduate-student projects, etc. Second, the judgment of what constitutes a replication attempt may vary between the coders, as well as with those with more specific expertise on the topic.

In the second part of our data collection attempt to corroborate our measures, assess economists’ awareness of replications, and get a larger sense for how many replications might be out there. To these ends, we surveyed two sets of authors. First, we sent personal emails to one author from each *Volume Sample* paper. Second, we sent a link to a survey for authors of papers in the *Citing Sample*.<sup>7,8</sup>

We elicited the beliefs of authors in both samples over the number of replications for the relevant volume paper that are (i) publications, (ii) working papers, and

(iii) projects never meant to be published.<sup>9</sup> Here our language in the elicitation was purposefully broad asking for the number of papers that “report a result that speaks directly to whether or not your paper’s main hypothesis is true.” For each response, we also asked about their confidence: “Do you think this number is pretty close, or is it more of a complete guess?” In total, one author from 36 of the 70 *Volume Sample* papers responded, and 58 of the 226 *Citing Sample* authors surveyed completed it.

### Survey Results

Overall volume authors were not sure on how many replications of their work had occurred. For the 25 *Volume Sample* authors that reported their confidence, just over half (13) were sure on their responses. Examining the 13 authors who were confident in their response, we do find substantial concurrence with our manual coding. All eight papers where a volume sample author pointed to one or more published works which have one or more replications or extensions coded in our data (5 as replications, 3 as extensions). For the remaining five papers where a volume author was sure there had been no replications, four were coded as having no replications.<sup>10</sup> In ad-

<sup>7</sup>Data collection for authors in the *Volume Sample* was more informal, with clarifying questions answered over email, and some qualitative responses reported. Data collection from the *Citing Sample* did not have as much two-way communication, and required numerical responses.

<sup>8</sup>In particular, we sent surveys to one author on each citing paper (with independent authors from the volume paper) that: (i) was coded as a replication; (ii) was coded as an extension; and (iii) cited the volume paper the most (and at least twice) but was coded neither as a replication nor an extension (95, counting ties). The goal of these criteria were to narrow the sample to authors who knew the volume paper well, but who were independent of the original.

<sup>9</sup>Authors in the *Volume Sample* were also asked how many of the working papers they believed would be published at some point.

<sup>10</sup>The only stand-out was explained within the email by a Volume respondent using a stricter definition of replication.

dition in all five cases where a volume author pointed to specific published work in our *Citing Sample*, the referenced work was always coded as either a replication or an extension.

Authors who were guessing on the number of published replications were in general more optimistic than those who were not. For the 11 volume authors who report a guess, only one guesses that their paper has no published replications; in contrast for the 13 authors who were sure in their numbers, five report no published replications.

This uncertainty on the volume authors part is reflected in the citing authors' responses. On a 0 (complete guess) to 100 (very sure) percentage scale, citing-sample experts report average confidence of 35 percent for published replication attempts, 27 percent for working papers, and 15 percent for informal projects.<sup>11</sup> Though this sample was hand-selected to be experts on this specific topic, there seems to be very little confidence in knowledge about replications.

The authors of citing papers estimate high rates of replication. The median belief on published replication attempts is two replications per volume paper, with almost three-quarters reporting that the paper they cited had at least one published replication. That rate increases to 83 and 78 percent for working papers and informal projects, respectively.

The surveyed rates of replication are substantially higher than our coded sample. This difference could reflect the narrowness of our coding—only using top-200 Economics papers—or differences in our subjective judgment of what constitutes a replication. However, considering the authors' beliefs to be the authoritative measure has to come with the qualification that the authors admit to being uncertain, and any mistakes will be biased towards more replications than zero.

<sup>11</sup>They are generally on the unsure side of the scale is meaningful: 74 percent report a confidence of 50 or below for their response regarding published replication work. That proportion increases to 84 percent for confidence in their beliefs of working paper replications and 97 percent for informal projects.

## Conclusion

Examining well-published papers and surveying experts in the specific topic, we find no general confidence in how many replication attempts exist. As a contribution towards shedding light on this uncertainty, the estimates from our coding exercise suggest that a majority of the very well-published papers in Economics are not being replicated at all—though well-published and well-cited works are being replicated at much higher rates.

There are reasons to suspect that the true rate of replication might be higher (or indeed lower) than the proportion we estimate. However, the measurements in our paper reflect very practical numbers: what economists believe, and what we can find through a search of the literature.

## REFERENCES

- Clemens, Michael A.** forthcoming. "The meaning of failed replications: A review and proposal." *Journal of Economic Surveys*.
- Hamermesh, Daniel S.** 2017. "Replication in Labor Economics: Evidence from Data and What It Suggests." *American Economic Review*, 107.
- Sunkhtankar, Sandip.** 2017. "Replications in Development." *American Economic Review: Papers & Proceedings*.