

Replicability of one journal: a comprehensive view

Sylvérie Herbert, Hautahi Kingi, Flavio Stanchi, Lars Vilhuber

2018-10-12

Abstract

Replication, reproduction, and falsification of published articles are an important part of the scientific endeavor, and has been widely discussed. The lessons learned from earlier replication exercises such as (Dewald et al, 1986) lead several journals to implement code and data depository requirements. These should have lead to improved reproducibility of journal archives, since a good part of the failure to replicate is due to poor availability of replication materials (McCullough et al, 2006). We set out to test the latter assumption. We assessed all articles published in the **American Economic Journal: Applied Economics** (AEJ:AE) as to the reproducibility of their computational results, and tested the reproducibility for all those with some data available.

We choose AEJ:AE primarily for two reasons. First, because of the empirical nature of its articles and its policy of publishing papers “only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication,” we expect that nearly all articles have some empirical component. Second, while other journals may also have theoretical or more complex empirical papers, we wanted articles to be reproducible by the average undergraduate student. Nothing in the methodology used in this paper is specific to that journal. We have extended the analysis to other AEA journals, targeting recent years of the **American Economic Journal: Macroeconomics** (AEJ:Macro), and **American Economic Journal: Microeconomics** (AEJ:Micro), though to date our efforts have less than complete coverage.

To date (the work is ongoing), we have assessed 686 articles. 154 papers don’t provide enough data to attempt a reproduction. We ultimately have verified the reproducibility of 401 papers.

In the paper, we describe in more detail characteristics of the papers that affect the reproducibility tests. For instance, papers with many analysis data sets, or with analysis data that is incompletely described, may pose challenges to undergraduates but not to experts in the field. We disentangle some of the possible causes that are correlated with partial or complete failure to replicate results. Finally, we look at the impact on the citations papers receive, and how those might be correlated with the difficulty to replicate or the failure to replicate.