



On the Need for a Replication Journal

Christian Zimmermann

Working Paper 2015-016A
<http://research.stlouisfed.org/wp/2015/2015-016.pdf>

August 2015

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

On the Need for a Replication Journal*

Christian Zimmermann
Federal Reserve Bank of St. Louis[†]

August 7, 2015

Abstract

There is very little replication of research in economics, particularly compared with other sciences. This paper argues that there is a dire need for studies that replicate research, that their scarcity is due to poor or negative rewards for replicators, and that this could be improved with a journal that exclusively publishes replication studies. I then discuss how such a journal could be organized, in particular in the face of some negative rewards some replication studies may elicit.

*This work benefitted from discussions with the participants of the Workshop “The Future of Scholarly Communication” at the Leibniz Information Centre for Economics in Hamburg. Comments from Graham A. Davis, Jan Höfler, Zacharias Maniadis, B.D. McCullough, Dan Newlon, Robert Reed are also acknowledged.

[†]Email: zimmermann@stlouisfed.org. The views expressed are those of individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

1 Introduction

In many fields of science, there is a strong tradition of replicating the results from previously published studies and then publishing the findings. However, such replication is virtually absent in economics, at least when publication is considered. There is no less need for replication, as economists consider that their research has an impact on our understanding of society and on public policies. Ill-informed policy could have disastrous results.

I define replication in the following way. In a strict sense, it is repeating a previous study to verify its validity (verification). This may be accomplished with the original data and computer code, but could also be performed with a resampling from the original data (reproduction) or with new data or computer code similar in scope to that in the original study (robustness). A broader definition involves using new data of different scope (different period, geography, or selection), a different solution or approximation method, or a new statistical approach (extension).¹

Despite the length of articles in economics, they cannot contain all the details of the work that was performed. When working with data, one often excludes some observations or otherwise manipulates the dataset to "clean" it in ways that may have an impact on results. In fact, this cleaning sometimes may have the purpose of obtaining statistically significant results, as Brodeur, Lé, Sangnier, and Zylberberg (2013) and Franco, Malhotra, and Simonovits (2014) may indicate, as published p-values have odd concentrations around critical values. The same applies to different identification strategies, sampling, control variables, or even datasets. The absence of such details, along with the reluctance to provide datasets and code to readers or referees, questions the full credibility of studies — hence the need for replication. Even when the data are provided, the precise handling of it matters, as the recent episode with Reinhart and Rogoff (2010) has shown (see Herndon, Ash and Pollin (2014) for details).

This concept is not limited to empirical studies. While economic theory has the tradition of providing full proofs in the paper (and referees are supposed to verify them), studies based on simulations may be influenced by the way approximations are taken, the choice of the solution method, the precision that has been applied, or even which software or software version is used.² In fact, many scientific contributions nowadays are in some ways encoded in software.

A new area where concerns about replicability have arisen is experimental economics. Researchers generate their own data and interpret it themselves. The collection of the data may be influenced by small details in the setup of the experiment that may not be recognized even if the documentation is extensive.³

Research without appendices cannot do justice to all the details of the research that has been carried out. In a sense, publication is only advertising research results, and the nitty-gritty details should be found in the actual data,

¹These definitions largely coincide with those from Clemens (2015), which discusses various aspects of replication.

²This applies of course also to empirical studies.

³The same could apply to other fields of economics, such as neuroeconomics.

computer code and even the notes of the researcher. Journal editors and referees do not receive this kind of detail when evaluating whether a manuscript is worthy publication. One gets a sense that they deem replication to be a post-publication question rather than a pre-publication one, which is certainly the case in other sciences. Why, then, is replication most often overlooked or absent in economics?

This lack of replication should not be due to the fact that the results are not important. It may be because data collection is costly, access to superior computing facilities is cumbersome, some datasets cannot be released to protect confidentiality, or cleaning datasets and programming or running simulations takes a long time. The relative dearth and small size of research grants in economics probably contributes to this lack of resources dedicated to replication. But we want to argue that the main issue is one of culture. There is no reward in replication, in particular because the vast majority of journals are not interested in them. It is a prisoner's dilemma. In the race for high impact factors, the first journal that would allow the publication of replications (and hence the occasional non-replication) devalues its content with respect to the other journals: replications are less citable, and non-replications contradict published content, which can cause ill-will. Journals are reluctant even to provide a separate section in which to publish replications.

In addition, few economists make replication easier by providing data or computer code, making the hurdle for the replicator even higher. Again, the incentives are not well aligned. Having developed or cleaned a dataset, the researcher wants to milk it for more publications. Opening the dataset to the public risks being preempted by an "opportunist."⁴ The same applies to computer code, and in both cases it opens the researcher to criticism about procedures. It is much easier to hide all that. It does not affect credibility that much, as nearly everyone proceeds in the same way. And only few journals force the hand of the authors by requiring datasets and code to be published.

The few systematic replication studies (i.e., not subject to some publication bias) that have been published show a bleak picture of the replicability of the economic literature. We will discuss some of these replication studies below, and I am sure readers know of more examples closer to their own field of study.

If the field of economics is to enhance or even retain its credibility, replication should be part of the literature. One way to achieve this is to actively promote replication. Existing journals could be more willing to accept replication studies, but it is clear reputational inertia is not providing the right incentives. While more and more replications are getting published, the pace of growth is slow. Hence, I suggest the creation of a journal entirely dedicated to replication.

In the following pages, I describe replication practices in other sciences. I then discuss various replication efforts in economics, to highlight both how few studies can be replicated and the difficulties in getting (non-)replications

⁴Although this can easily be prevented by releasing the dataset only at the time of the publication of the article.

published. I finally argue that we need a journal dedicated to replication in economics and suggest ways in which such a journal could operate despite the profession's reluctance.

2 Replication in other sciences

A large part of the body of literature in hard and medical sciences is dedicated to experiments.⁵ In many cases, the samples under study are much smaller than what economists would tolerate. However, there is a strong tradition to try to replicate new results, and the publication process is quick enough to get such replications published soon after the original study. If subsequent studies fail to replicate, it is not uncommon for journal editors to publish corrections and more commonly retractions of the original study. In some cases, even the original authors ask for a retraction. A retraction is useful because it prevents, in principle, other authors from citing a flawed study.

In some sense, the hurdle that a study has to pass to be published in hard and medical sciences is significantly lower because it is understood that the peer review process does not deal with replication. Basically, it vets the seriousness, the plausibility, and the relevance of the study. Replication comes post-publication, and it is taken very seriously.

In humanities and social sciences other than economics, replication studies are extremely rare. For humanities, the nature of the studies is simply not amenable to replication: few datasets are used and experimentation is unknown. Social sciences have no excuses. Political sciences are making an effort. *Political Analysis*, a top journal in the field, has a section dedicated to replication. The *Political Science Replication Archive* has recently started. Replication is extremely difficult in anthropology, as case studies and narratives are prevalent. Still, Syracuse University is now hosting a Qualitative Data Repository. Finally, as sociology is becoming increasingly quantitative, replication studies seem likely to follow.

3 Replication in economics

The publication process in economics is quite distinctive. Papers are much longer and detailed, and the peer review process can take years with substantial revisions requested by referees and editors. One would think that with this level of scrutiny, referees or editors would take a look at the code or the data, or try to replicate the study. None of this is happening. In fact, I know of no journal that would require authors to deliver anything beyond a manuscript at submission. And journals do not have the means to verify replicability themselves, as software is expensive and replication does take time. Under such circumstances, one

⁵The first to stress replicability was Boyle (1660) who provided extremely detailed descriptions of his experiments so that others could try to reproduce them and be convinced. With current publication technology, this approach should not be a problem.

would think that replication would happen after publication, as it is in hard and medical sciences. But there is hardly any replication, and attempts to publish replications are made very difficult, as some examples below show.

3.1 Replication attempts

Here, I want to discuss some systematic attempts at replicating previous studies. These do not focus on a particular study, but rather on a class of studies.

It should be noted that replication is not only important because it checks whether the ivory tower has done the right thing. It is policy relevant and errors, “short-cuts,” or deliberate mistakes can matter, as forcefully discussed by McCullough and McKittrick (2009).

Dewald, Thursby and Anderson (1986) performed an interesting experiment with articles from the *Journal of Money, Credit and Banking*: They asked 62 authors of already-published articles to supply code and data: 18 ignored repeated requests, 20 refused, 2 cited confidentiality, and 22 responded. Another 92 among those with papers under review or accepted received a similar request, after it was made clear at submission such a request would be made: 17 never responded, 7 declined, 68 supplied something. 54 were analyzed, only two of which were deemed to be sufficiently complete and documented to warrant replication. The lesson from this early exercise is that making it clear that supplying materials will be required helps,⁶ but it is not sufficient. The *JMCB* subsequently made code and data mandatory, which McCullough, McGeary and Harrison (2006) took as opportunity. Of 186 candidate articles, only 69 had something deposited. Of the 62 for which software was available for replication, only 14 could actually be replicated. The archival policy was visibly not enforced, and it was not checked that authors actually were supplying relevant or complete materials. McCullough, McGeary and Harrison (2008) turn their attention to the Federal Reserve Bank of St. Louis *Review*; they were somewhat more successful here likely because enforcement was better for this publication. Authors need to make an effort to facilitate the reproducibility of their study, and only very few do it. My experience as the “code and data archivist” of the *Review of Economic Dynamics* shows that the situation has considerably improved. It is, however, a requirement for this journal to provide materials before an article is accepted, and I will pester authors until the documentation is appropriate. But not every editorial team may be willing to make this effort.⁷

To highlight how depositing code and data, even if well documented, is not sufficient, consider the point raised by Yalta and Yalta (2012). They run accuracy tests on various econometrics software, looking at different versions of the same software. They find that errors remain uncorrected for many years in commercial software, while they are quickly fixed in open-access software. What this means is that with the same dataset and the same code, results may vary

⁶McCullough and Vinod (2003) find dismal response rates from authors with journals that adopted a policy requiring authors to supply data and code upon request.

⁷Note that our download statistics indicate that replication code and data for *RED* are wildly popular.

across software versions, with older versions more likely to have errors that are no fault of the authors.

The problem is not limited to empirical studies. Kleiber and Zeileis (2011) show that code for simulation studies is similarly lacking. Instructions almost never specify which software, which version, or which seed is used for random number generators. In many simulation papers, the details of the data-generating process are crucial.

3.2 (Non-)Publication of replication

Attempts to publish replications of single studies have been much less successful, with mounting anecdotal evidence that replicated authors are hindering publication when replication does not corroborate the original results. This discussion is based on feedback from several replicators, some of which have requested anonymity.⁸ This is neither an exhaustive nor a representative list, but it highlights some of the issues replications face in economics. All authors are purposefully referred to as males.

3.2.1 Case 1

Author A publishes a path-breaking empirical study in a top journal that greatly contributes to his reputation. Author B challenges the results and shows that the article is not path-breaking at all. Author B openly challenges Author A and a public feud ensues. Author B does not manage to publish his study and also suspects that Author A is using his influence to block Author B's other publications.

3.2.2 Case 2

Author C is interested in a particular topic and finds that the literature on the topic is motivated by one empirical study that performs a rather simple instrumental variable estimation. Empirical methods have progressed since this seminal study was published, so he is interested in replicating the original study as a first step and plans to apply newer methods as part of a second step. He asks original author (Author D) for the data. Author D claims the data have been lost.

Author C then proceeds to piece the data together, trying to use historical series as they were available at the time of the original study. He manages to replicate the original results (up to a few typos) after dropping some specific observations. He notices, however, that the instrument is statistically very weak. After an extensive search, he finds a satisfactory instrument and all results of the original study become non-significant.

Author C submits his paper to several journals, each time rejected on the basis of one very negative referee report. It is strongly suspected that Author D is the referee. Author D's reputation is largely built on the original study, thus it

⁸It is also not the goal of this paper to name and shame obstructors.

is understandable that he is trying to protect his reputation. Editors keep asking Author D to referee the submitted paper, despite pleas from Author C to bypass him. But editors do not discount his reports. As Author C keeps submitting to more journals, eventually this referee tries to persuade Author C modify the paper so that the original study remains in a good light. Interestingly, the referee now uses the original dataset in his motivation. Some middle ground is eventually reached and the replication is now published, although to a large extent it is ignored because its message has been muted.

3.2.3 Case 3

Author E publishes several papers in top economics journals where simulations form an essential part of the results. The computer code for each of the articles is based on a common core of code. As is later realized, that code contains some errors that make some results impossible to replicate with new code. Author F, who was working on an extension of one of the original studies abandons the work after failing to replicate the original. Author G, seeking to extend another study, also notices a failure to replicate, but forges on and submits to a journal. Its editor accepts the paper after hearing from yet another author that the original study is flawed. This fact is, however, not noted explicitly in the published article, and neither is it noted anywhere else in the literature.

3.2.4 Case 4

Author H publishes an empirical study that resolves a puzzle. Author I notices that there is a decisive flaw in the paper that invalidates the results. Author I submits the analysis to the same journal but is rejected. Author H is allowed to publish a correction (but not a retraction) in the same journal. Author I only manages to publish the analysis in a minor journal.

3.2.5 Case 5

Author J submits a study for publication that uses proprietary data obtained from several companies in an industry group. Author K, who is financed by a rival industry group, has serious doubts about the results and asks for the code and data. J agrees to provide the data, then objects, citing the confidential agreement, which he never produces because he claims it is also confidential. The journal editor accepts Author J's paper and does not respond to Author K's concerns, for example by requiring the peer reviewers to look at the data. A paper by Author K, who is failing to replicate the results based on publicly available data, is rejected by the journal, with one referee arguing for publication, the other against. The latter turned out to be Author J. While there may be some concern about bias due to the financing of both authors, the lack of openness means that we have not progressed in understanding the discussed issue.

3.2.6 Case 6

This case is documented in Davis (2007), who obtains data about another study and fails to replicate it. He sends his analysis to the journal where the original study is published. It is rejected eight months later based on a single, one-paragraph referee report that acknowledges the original paper has errors. A request to send to a second referee remained unanswered. The replicator then sends the paper to four other journals, one rejecting the replication for being out of scope, the other three basing their decisions on one or two referees reports. All of the latter three submissions received the exact same referee report as in the very first submission. The last and fifth journal accepted the paper.

The previous examples constitute a biased sample: In some way their history became public. There are others we do not know about, and then the replication studies that never happened because of the culture that does not encourage them. Indeed, even for this biased list of examples, we see that many authors — especially well-established ones⁹ — whose results are not successfully replicated are quite successful in hindering the publication of a replication study, in large part with the complicity of the editors. And there may be many others that have been kept hidden, not counting all the replication studies that have never been made public.

3.3 Replication databases

With funding from the Institute for New Economic Thinking, the *Economics Replication Wiki* at Universität Göttingen (Germany) has recently started an index of replicated studies (including information on data, methods, and software), as well as publishing a priority list of studies that should be replicated. This ongoing effort demonstrates that there is a non-negligible number of replication studies in economics, some even published. It is, however, nowhere near the number found in other sciences. The Wiki also demonstrates that there is an alarmingly large proportion of publications in top journal, many of them seminal, for which no replication is known.

The *American Economic Association* has recently sponsored a registry of randomized control trials. It asks the researcher to deposit their research plan, including the experimental design, before they go to the field for data collection. The goal is to provide a platform for principal investigators to commit to a particular design and thus avoid the temptation of datamining for significant results. With regard to replication, a positive externality of this registry is that it allows tracking of all relevant studies, not just the ones with positive results that were published. However, as Coffman and Niederle (2015) argue, such a repository is useful only for studies that are very difficult to replicate: a costly one-of-a-kind field experiment will never get a robustness check, let alone

⁹Note that all the original studies mentioned above except for the last are among the top 1% items listed on RePEc (<http://ideas.repec.org/top/#items>).

a replication. A pre-analysis plan makes the results more credible. In the end, it is not clear it makes replication more desirable or likely.

The *3ie International Initiative for Impact Evaluation* has opened a replication paper series that invites replication studies of development impact evaluations. At this point, four studies have been published, and replicated authors have the opportunity to respond. The studies are typically funded by 3ie or conducted in-house and the opinion of one external referee is solicited. This does not equate them with journal articles, though. As Ozler (2015) highlights, the fact that *3ie* funds these replications makes them impossible to reject. All should eventually be published, regardless of their merit.

The *Behavioral Economics Replication Project* has taken a very particular approach. The project leaders identified 18 laboratory experiments published from 2011 to 2014 in the *American Economic Review* and the *Quarterly Journal of Economics*. All of them are going to be replicated as part of a meta-experiment: This was preceded by a market that would price the probability the main result of each paper would be successfully replicated. Results are pending, and it will be interesting to see whether this initiative will extend beyond the first set of experiments.

3.4 Rewards for replicators

As noted, very few journals publish replications. As inventoried by Duvendack, Palmer-Jones and Reed (2015), six journals account for 60% of all replications they found (and only 24 of the 333 top journals even have a data and code repository policy). Someone wanting to publish a replication study will find only one or two possible outlets once excluding journals that are out of the field. One possible outlet is the *Journal of Applied Econometrics*, which publishes by far the most replications, but it is limited to empirical studies initially published in the very top journals and it has limited space.

A potential candidate is *Econ Journal Watch*, which is an online-only, open-access journal whose goal is to “watch the journals for inappropriate assumption, weak chains of argument, phony claims of relevance, omissions of pertinent truths, and irreducibility (*EJW* also publishes replications).” Hence, this outlet should be a good vehicle for replication studies. After 11 years, relatively few have been published, however. The fact that a strong emphasis is put on replication failures, and the fact that the journal does not hesitate to be confrontational, has probably prevented *Econ Journal Watch* from becoming an outlet that replicating authors vie for.

The new *Journal of the Economic Science Association* promises to be explicitly receptive of replication studies. The scope of the journal, however, is limited to experimental economics. Other journals have occasional calls for replications in a specific area (*Public Finance Review*, *Empirical Economics*, *Applied Economics*), but replications are still not treated as ordinary studies. And note that replications of simulations have been completely absent in this section.

As far as I know, there is no other journal that would be a natural outlet

for replication studies. And the journals above cannot publish all replication studies, for obvious reasons. Other journals publish very few such studies. Why? Editors are simply not interested. A successful replication is not likely to attract citations, as there is nothing new to report. The vast majority of editors seek studies that are path-breaking in some way or add a new twist to an existing result. Merely confirming a result does not fit the bill. If a paper fails to replicate, there is a contribution to the literature that editors should be seeking. However, and as discussed above, this is potentially inviting trouble. In addition, a failed replication comes with uncertainty as to who is at fault, the default assumption seemingly being that the replicator is not correct. An editor is then particularly hesitant to accept a replication study.

4 A proposal: a journal dedicated to replication in economics

As mentioned, a journal dedicated to replication could provide the necessary rewards for replication. For this to happen, the journal needs to be credible and build a good reputation. Necessary conditions for this are a good editorial team, strong procedures, and convincing selection criteria.¹⁰

4.1 Why such a journal is needed

Given the lack of current rewards for replication studies, we need to create some if we want to see more of them pursued. The best way to do this is to create a journal dedicated to them. In economics, many appointment, tenure, and promotion decisions are based on the publication records of the researcher, who allocates her efforts accordingly. If there is a credible peer-reviewed journal for replications, replications are more likely to happen.

Why does replication matter? As detailed above, there is currently no culture in the refereeing process that includes checking and replicating results, despite the absurd time peer-review is taking in economics. We need the post-publication replication tradition that other sciences have. This is important because policy implications of results of papers that cannot be replicated can be disastrous. McCullough and McKittrick (2009) cite a few spectacular examples.

It should be noted that the fact that some journals make the code and data of their articles available is not sufficient. As noted above, documentation is most of the time seriously lacking, making replication difficult. Also, replication does not necessarily mean redoing the work with the same code and the same data. Checking for robustness to other datasets or numerical procedures is important. For example, for randomized experiments, studies are performed

¹⁰For another recent proposal for a replication journal in experimental economics, see Coffman and Niederle (2015). Note that there is a *Journal of Negative Results in Biomedicine* that only publishes insignificant results, realizing that they are important but unpublishable elsewhere.

in a particular setting that may or may not be generalizable. But, as Duflo, Glennester and Kremer (2008) point out, there are no incentives to perform replications in other setups. Or Johnson, Larson, Papageorgiou and Subramian (2013) show that different versions of the Penn World Tables, which are widely used for cross-country regressions, can yield wildly different results for many widely cited studies.

4.2 How it could be organized

I see four general areas that require replication studies: empirical microeconomic studies, empirical macroeconomic studies, simulation studies, and experiments. I briefly describe each area.

Empirical microeconomic studies use data at the individual, household, or firm level. These are cross-sections and sometimes panels. They have a large number of observations and may be extracted from surveys, censuses or administrative databases.

Empirical macroeconomic studies pertain to countries or other geographic aggregations. A majority of them use time series, although there may be some cross-country regressions or panels.

Simulation studies solve artificial economies drawn from a theoretical model that may have been calibrated or estimated. Such economies typically solve for an equilibrium and then run counterfactuals, for example to look for the impact of a policy. Simulations are also frequently used in econometric theory.

Experiments come in two kinds: large randomized interventions mainly in labor and development economics and small experiments with paid participants in a laboratory setting usually to test fundamentals of microeconomic theory.

I see each of those four areas covered by an individual editor, who solicits referee reports and then submits a letter to the other three editors based on her own reading and the reports. This letter should remain factual and not indicate what the decision should be. The other three editors then vote anonymously on whether the replication study should be accepted for publication, a revision should be requested, or it should be rejected. A managing editor allocates the papers to editors, conducts the votes, and corresponds with authors. He does not participate in votes.

It is rather unusual that the final decision be taken by editors outside the field. It is purposeful. Given the current animosity toward replication studies and among involved authors,¹¹ it may be best to avoid making the editor an outcast within her group of peers. Adding a managing editor could add a further layer of protection.

What would the criteria for inclusion in this journal be? I would see the following main ones, in order of importance. Note that making available code and/or data with full documentation is a requirement. This makes studies more credible in general and replications more useful.

¹¹Maniadis, Tufano and List (forthcoming) also highlight this unfortunate aspect of current replication attempts.

1. The importance of the replicated study. This could be in terms of importance in the literature (citations) or in terms of policy relevance and implications.
2. Improvements to the original study. Replicated studies can be improved with more robust estimators, larger samples, better identification, better solution methods, or better controlled experiments.
3. Lack of previous replication. Quite obviously, the more a study has been replicated, the smaller the marginal contribution of an additional replication, all else being equal.
4. Failure of replicating the original study. While a replication success is worthwhile, special attention needs to be addressed toward failures, as they may have major implications on the literature or policy. The fact that this is not a top requirement highlights that replication success should also be published, though perhaps devoting comparatively less space to it. In fact, it may be important that there is no publication bias in this respect, as Coffman and Niederle (2015) calculate that roughly five replications are sufficient for a result to be convincing if there is little such bias. Without such a bias, Maniadis, Tufano and List (2014) think four replications are sufficient.
5. Code and/or data not previously made available. This makes the need of replication all the more necessary. Note that even if code and/or data are available, replications fail, for example because the code is buggy or the dataset is incomplete. This should be equated to code and/or data not available.
6. Difficulty of replication. An obviously easy replication should have less merit. Difficulty can come from access to the data, size of the data-cleaning or programming task, complexity of computations, or necessary cost of experimentation.

One should also avoid some easy pitfalls. One is turning replication exercises into witch hunts. By this, I mean trying to find something wrong in every paper. One example of this is Dercon (2015) showing frustration that replicators focus on the few specifications that do not replicate the paper, specifications he clearly documented not to be of interest in the working paper version. Ozler (2014) makes a similar point. One way to avoid this pitfall is to consider only strict replications (verifications) or, in the case of experimental economics, replications with newly gathered data.

5 Conclusion

Natural and life sciences have a strong traditions of replicating previous studies, publishing the results, and retracting invalidated studies. Social sciences and

humanities do not. This is especially necessary for economics, first because the field is data intensive, second because economists like to think their work has strong and credible policy implications, third because economists also like to think that their field strives to achieve the rigor of natural sciences. Animosity toward replication and little reward for replication limits publication of such studies. A journal dedicated to replication in economics could provide the right rewards if it has strong editorial backing and rigorous procedures. Even dedicating space in existing journals to allow for some short notes on replication attempts would help.

A replication journal in isolation is, however, not likely to be sufficient to create a healthy replication culture in economics. Mandating that data and code be made available is an important part of such a culture, and there has been movement in that direction as several top journals and funding agencies now require this. Adopting better citation practices for datasets and code would help. Pre-registering experimental and large-scale empirical studies is another step. Requiring successful replications of one's work for tenure or promotion or generally to establish research quality could also provide interesting incentives (Maniadis, Tufano and List, forthcoming).

6 References

Boyle, Robert, 1660. "New Experiments Physico-Mechanicall, Touching the Spring of the Air, and its Effects (Made, for the Most Part, in a New Pneumatical Engine) Written by Way of Letter to the Right Honorable Charles Lord Vicount of Dungarvan, Eldest Son to the Earl of Corke." Oxford: H. Hall.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg, 2013. "Star Wars: The Empirics Strike Back," IZA discussion paper 7268.

Clemens, Michael, 2015. "The Meaning of Failed Replications: A Review and Proposal," Center for Global Development working paper 399.

Coffman, Lucas, and Muriel Niederle, 2015. "Pre Analysis Plans Help (only) when Replications are Infeasible," mimeo to be presented at the Econometric Society World Congress in Montreal.

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott and Tassew Woldehanna, 2015. "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages: Response to William Bowser," 3ie International Initiative for Impact Evaluation, response to Replication Paper 4.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson, 1986. "Replication in empirical economics: the Journal of Money, Credit and Banking project," *American Economic Review*, vol. 76(4), pages 587–603.

Duflo, Esther, Rachel Glennester, and Michael Kremer, 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*, vol. 4, T. Paul Schultz and John Strauss (eds.), pages 3895–3962.

Duvendack, Maren, Richard W. Palmer-Jones, and W. Robert Reed, 2015.

“Replication in Economics. A Progress Report,” *Econ Journal Watch*, vol. (12)(2), pages 164–191.

Franco, Annie, Neil Malhotra, and Gabor Simonovits, 2014. “Publication Bias in the Social Sciences: Unlocking the File Drawer,” *Science*, vol. 345(6203), pages 1502–1505.

Davis, Graham A., 2007 “Reflections of a First-Time Replicator,” *Journal of Economic and Social Measurement*, vol 32, pages 263–266.

Herndon, Thomas, Michael Ash, and Robert Pollin, 2014. “Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff,” *Cambridge Journal of Economics*, vol. 38(2), pages 257–279.

Höfler, Jan, 2014. “Teaching Replication in Quantitative Empirical Economics,” mimeo presented at the European Meeting of the Econometric Society in Toulouse.

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian, 2013. “Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates,” *Journal of Monetary Economics*, vol. 60(2), pages 255–274.

Kleiber, Christian, and Achim Zeileis, 2011. “Reproducible Econometric Studies,” Working Papers in Economics and Statistics 2011-02, University of Innsbruck.

Maniadis, Zacharias, Fabio Tufano, and John A. List, 2014. “One Swallow Doesn’t Make a Summer: New Evidence on Anchoring Effects,” *American Economic Review*, vol. 104(1), pages 277–290.

Maniadis, Zacharias, Fabio Tufano, and John A. List, forthcoming. “How to Make Experimental Economics Research More Reproducible: Lessons from Other Disciplines and a New Proposal,” *Research in Experimental Economics*.

McCullough, B. D., 2009. “Open Access Economics Journals and the Market for Reproducible Economic Research,” *Economic Analysis and Policy*, vol. 39(1), pages 117–126.

McCullough, B.D., Kerry Anne McGeary, and Teresa D. Harrison, 2006. “Lessons from the JMCB archive.” *Journal of Money, Credit and Banking*, vol. 38(4), pages 1093–1107.

McCullough, B.D., Kerry Anne McGeary, and Teresa D. Harrison, 2008. “Do economics journal archives promote replicable research?,” *Canadian Journal of Economics*, vol. 41(4), pages 1406–1420.

McCullough, B. D., and Ross R. McKittrick, 2009. “Check the Numbers: The Case for Due Diligence in Policy Formation.” *Studies in Risk & Regulation*, February, The Fraser Institute.

McCullough, B.D., and H.D. Vinod, 2003. “Verifying the solution from a nonlinear solver: a case study,” *American Economic Review*, vol. 93(3), pages 873–892.

Ozler, Berk, 2014. “How Scientific are Scientific Replications?,” Development Impact Blog, World Bank, 15 October 2015, <http://blogs.worldbank.org/impactevaluations/how-scientific-are-scientific-replications>

Ozler, Berk, 2015. “Response to Brown and Wood’s “How Scientific Are Scientific Replications? A Response””, Development Impact Blog, World Bank, 25

January 2015, <http://blogs.worldbank.org/impac evaluations/response-brown-and-woods-how-scientific-are-scientific-replications-response>

Reinhart, Carmen M., and Kenneth S. Rogoff, 2010. "Growth in a Time of Debt," *American Economic Review*, vol. 100(2), pages 573–78.

Yalta, A. Talha, and A. Yasemin Yalta, 2012. "Should Economists Use Open Source Software for Doing Research?," Hacettepe University Department of Economics Working Papers 20127.