

Editorial: An Opinionated FAQ

Ivo Welch

ivo.welch@ucla.edu

I want to share many difficult questions for which I do not know the answers. To help the conversation along, I am also providing my own current humble opinions (IMHO), subject to change.

1. Should journals publish only unambiguous correct papers?

– No. If they did, they would not publish anything. Complex computer programs are rarely bug-free. Data usually has some errors. Textual descriptions are always incomplete. Methods rely on models that are approximate. Plus, authors can commit different types of errors in their analyses, with wide gray zones between: (a) honest data and code errors; (b) cherry-picked specifications and results; (c) priors that led to stopping investigation when good results seem to have fit the hypothesis; and (d) outright academic fraud. Hopefully, problems and errors are few and innocuous, and occur primarily in “edge” cases that have little material effect on the results and inference.

Unfortunately, Darwinian “continuation selection by author” and “publication selection by referee” favor papers with stronger and more “interesting” results. Thus, the frequency of serious problems may well be higher among published than unpublished papers.

2. Should it be sufficient that published papers are reproducible *in principle*?

– No. In my opinion, papers have to have been reproduced *in actuality* by an independent and skeptical *devil’s advocate* in order to be considered reliable. Until then, even published results should only be considered “tentative.”

3. Can independent researchers *reproduce* (not just *replicate*) published results without the original data and code?

– Rarely. The reexaminations in this volume show that even contemporaneous reproduction attempts could not agree with one another.

4. If a published paper was wrong, and data and code are not available, can the academic record be corrected?

– No. Worse, there can be no reconciliation between the original published findings and any new findings, leaving authors and readers eternally unclear.

5. Should journals publish papers if the authors do not make data and code readily available?

– No. Without data and code, the danger of errors, both innocuous and malicious, is too high. If there is a consequential error, it will likely never be identified, and the record will likely never be corrected.

Moreover, data and code availability are a positive externality. They lower our collective cost of conducting research. It was acceptable for journals in the 20th century not to insist on researchers providing data and code. It should no longer be acceptable today.

6. Should journals publish papers with proprietary data or code that cannot be made accessible?

– This is a dilemma, but my answer is no. *Non-reproducible single occurrences are of no significance to science* (Popper, 1935). Even when replication is feasible, the current crisis in science has shown that results are not trustworthy until actually replicated. Unreplicable studies are likely to be even less reliable than unreplicated studies.

In special and rare cases, the journal, authors, and data providers could try to arrange for one or more analyses of the proprietary data by sanctioned (skeptical and independent) *devil's advocates* reproductions and replications.

If papers based on proprietary data are exempt from code and data requirements, then selection will shift authors towards work with more proprietary data—a counterproductive outcome.

There is a good place for proprietary-data studies. It is on unrefereed websites, like SSRN, which accept all comers without pretense of peer review, journal endorsement, or scientific replicability. Citations to such studies should be required to note the lack of independent verifiability.

7. Would a universal data and code requirement impose a cost on the profession?

– Only a modest one. It is true that preparing and cleaning up data and code will cost original researchers some time at the final publication stage (but also possibly help catch bugs). This is outweighed by the direct benefit of collective sharing. Code and data will become useful for iterative progress on important questions, and there will be fewer false uncorrected findings in the literature wasting everyone's time.

8. Should journals have obligations to themselves, their readers, or their authors?

– Ideally, journals should have obligations only to their readers and authors. In reality, editors often care about their journal impact metrics first, because these attract more readers and authors.

9. Should journals simply refuse to publish papers that lack skeptical replicability?

– Ideally, yes. Unfortunately, journals are in a competitive struggle for citations and impact, just like researchers. False surprising findings are often more visible than correct unsurprising ones. Such perverse competitive journal incentives can only be overcome collectively.

10. Should journals vet research papers that they are publishing?

– Yes. Today, our journals publish papers without ever fact-checking them. (Ironically, popular magazines do more than “scientific” journals.) “Surviving peer review” means that only editor and referees liked the paper and did not spot *described* obvious flaws. We are ripe pluckings for our own Sokal hoax.

Vetting can be accomplished without the journal reproducing the results. Instead, journals could set aside one-third of their pages to skeptical replications conducted by *devil’s advocates*. How much more trustworthy (and clearer) would our academic literature be if its accepted knowledge base would have had to survive such independent analyses?

11. Should journals have an obligation to publish critiques of papers that they have endorsed (by publishing them)?

– Yes. Presumably, the topic and paper were important enough. Why would the same journal not find corrections to be interesting?

12. Should journals publish retractions?

– Yes. The lack of any retractions in financial economics is not something to be proud of. It is not likely due to exceptional standards of care. Unfortunately, publishing retractions (and critiques) is rarely advantageous either to authors or to journal impact metrics.

13. What kinds of corrections and updates should journals publish?

– Ideally, all. Empirical findings should be like software, subject to regular releases, iterative progress, bug fixes, and updates. Less important updates could live on the website of the paper. More important updates deserve primary journal space. To be feasible, updates and fixes should strive to be short and to the point, so that the *devil’s advocate* can quickly confirm them. Ideally, they should be posted together with the original paper for improved saliency.

Such a process would prevent insidious *bad habits drift*, where errors that were at first innocuously improving significance, later establish precedence, accumulate over time, and eventually become convention. Unfortunately, an iterative corrective publication process is not on the horizon. The CFR is the next-best thing.

14. How can the journals encourage skeptical examinations?

– (1) They should commit never to ask the original authors to weigh in as anonymous referees. (2) They should commit to a number of pages dedicated to reproductions, replications, updates, and critiques. (3) They should commit to publishing at least one skeptical replication for each empirical original paper.

15. Should journals ask original authors to referee critiques?

– Yes and no: never *anonymously*. For gathering information, the original authors should be allowed to write a *non-anonymous* response, which should be forwarded (with named identities) to the real referees, who understand the conflict of interest. This is how the CFR handles critiques.

16. Do journals owe authors and the profession unconflicted referee reports?

– Yes, because this encourages more honest papers and especially critiques. The original author of a critiqued paper is usually informed but almost never objective. With acceptance rates of 1 in 10 or less, even one negative review is likely to bias the editor's decision towards rejecting.

17. Are conflicted referees a less serious problem than conflicted authors?

– Anonymous referee conflicts of interest are worse than author conflicts of interest, because such conflicts of interest remain anonymous, and they can shape academic literatures just like conflicted papers. (Critiqued authors are just the most obvious manifestation.) Having one's own research career built on the same or opposite findings is another important conflict of interest that should require disclosure to editor and authors. (Then it is fair game, however.) If honest, open agreement or disagreement is impossible, then such referees should excuse themselves from the process.

18. Have lax data and code requirements contributed to more amazing results?

– Probably yes. I have heard many researchers opine that research is expected to “put its best foot forward,” and they expect this to have happened when they are themselves referees. They infer that a more objective version of the results would likely have been less remarkable. (And, probably, they usually are!) In turn, over time, many researchers have become advocates rather than scientists in order to survive. All of us have struggled with the gray area of which data presentation is and which is not ethical and customary. Over time, submissions to our best journals had ever-more remarkable results. More boring results were no longer worthy. Their authors did not get tenure and disappeared.

19. What is the externality if journals do not commit to independent skeptical verification?

– The peer-evaluation academic rat-race forces researchers to prioritize quick and remarkable results over careful and boring results. And authors with incorrect results later become expert go-to referees, suppressing correct findings. Incorrect but surprising results often have higher survival probability in the refereeing process. The literature eventually stretches credibility.

20. Should we (authors) cite papers which do not make their data and code accessible?

– This is a difficult question, but my answer is no for more recent publications.¹ I have two reasons. First, not sharing data and code when similar papers have done so should raise a red flag about the results. Second, such papers have imposed a negative externality, in that they have poisoned the publication pipeline for similar papers whose authors would have been willing to share their data. Such papers do not deserve credit.

21. Do researchers have an ethical obligation to report when they cannot replicate/reproduce a paper?

– Ideally yes. Otherwise, how many other authors will have to suffer the same fate? However, if the original data and code is not available, the researchers cannot be sure that the original authors made the mistake rather than they themselves. Moreover, publication of negative results can be time-consuming, difficult, and upset the original authors.

22. Should researchers broadcast informally when they cannot replicate/reproduce a paper?

– No. Researchers should only do so if they post a paper that shows their work. We should not assassinate papers by insinuation.

23. Does a critique have to reject all findings in a literature?

– No. Quantity is no substitute for quality. Every paper (and sometimes every important claimed finding) can be critiqued in itself. A critique need not overturn each and every finding in a literature to be relevant.

24. Should our profession do more to encourage carefulness over speediness?

– Yes. Unfortunately, the rat-race effects of precedence reward “sloppiness and being first” over “carefulness and being second.” Data and code requirements can perhaps mitigate but not correct this problem, because it can help to uncover sloppiness.

¹If a CFR author decides not to cite a recent paper for which the data and code should be, but are not available, I will support this decision.

25. How should we (our profession and you) consider papers which have overstated their findings?

– This is a dilemma. When does the end justify the means? For example, if a paper rejected the null hypothesis, claiming an incorrect T -statistic of 3.0 instead of a correct T -statistic of 1.0, then the hypothesis still has an 85% probability (at its T of 1.0) of turning out to be correct in the end. Does this paper then deserve credit for the discovery? Or should the credit go to the first paper that, perhaps with more and newer data and better techniques, documented a correct publication-worthy T -statistic? Or is this newer paper already “scooped” and no longer deserves publication?

My remaining three questions are much easier:

26. How should code and data be archived and distributed?

– Storing data on one single storage medium is not good enough. Although hard disks and solid-state disks are more reliable today than they were 20 years ago, they still fail. Authors still need to retain multiple copies in multiple places, preferably with one copy on a near-indestructible M-DISC.

27. How should code and data using commercial data (such as CRSP and Compustat) be archived and distributed?

(1) For public data distribution: (a) detailed sample statistics (number and missing observations, multiple percentiles, etc.) for each series; and (b) a few complete example observations (including some extreme observations).

(2) For full reproductions, the author should keep a private copy of the full data set with which the results were obtained.

28. Should authors, journals, or associations retain publication data and code?

– All parties should retain it. Sadly, the CFR is not profitable and this management job exceeds our current capabilities. A better solution would be a central repository that contains code and data for many journals, linked to from each article’s publication web page. Such a repository should be maintained by our very profitable flagship *American Finance Association*.

Reference

Popper, K. 1935. “Zur Erkenntnistheorie der modernen Naturwissenschaft.” In: *Logik der Forschung*. Springer Verlag, p. 17.