The views and opinions expressed here are mine and are not necessarily those of the Board of Governors of the Federal Reserve System

Response to Referee 2 for "A Replication Recipe: List Your Ingredients Before You Start Cooking"

November 24th, 2017

Dear Referee 2,

Thank you for reading and commenting on my paper.

The special issue that I submitted this paper to was "The Practice of Replication" for Economics: The Open-Access, Open-Assessment E-Journal. A link to the call for papers can be found here.

In particular, requirements for the special issue are:

Contributors to the special issue will each select an influential economics article that has not previously been replicated, with each contributor selecting a unique article. Each paper will discuss how they would go about "replicating" their chosen article, and what criteria they would use to determine if the replication study "confirmed" or "disconfirmed" the original study... Note that papers submitted to this special issue will not actually do a replication.

Submitted papers will consist of four parts: (i) a general discussion of principles about how one should do a replication, (ii) an explanation of why the "candidate" paper was selected for replication, (iii) a replication plan that applies these principles to the "candidate" article, and (iv) a discussion of how to interpret the results of the replication (e.g., how does one know when the replication study "replicates" the original study).

The contributions to the special issue are intended to be short papers, approximately Economics Letters-length [about 2000 words].

As you suggested, I will add a link to the call for papers to the next draft.

For your remaining comments *(below, in italics).*

1. I fail to see how the statements made about replication attempts do not apply more generally to essentially all scholarly research projects. Consider the use of statements like

"Without prespecification, the amount of flowtime and budget that you could invest in a replication could grow uncontrollably" and "Your budget for a 'successful' replication is, most likely, less than that of the Bill & Melinda Gates Foundation". Replace the word 'replication' with 'study' and you still have equally true statements.

I should clarify these statements to indicate that the amount of flowtime and effort that you could invest to get a "successful" replication could go to infinity, as opposed to just doing the replication. At some point a researcher who is attempting a replication, after not seeing "success" for some amount of flowtime and effort, will have to declare that he/she was unable to replicate. What I argue in this paper is that the researcher should prespecify that amount of flowtime and effort.

While the amount of flowtime and effort you could spend on an original research study could also go to infinity, original research studies differ from replications in that the researcher doesn't know the results before undertaking the original research study (or, at least, shouldn't know). After spending flowtime and effort on an original research study, the researcher will have some type of result (or, at least be in the process of generating results), but whatever those results are they do not correspond to a hard definition of "success" like with a replication, because the researcher of an original research article, hopefully, is not trying to match a particular result.

2. Several arbitrary and/or unnecessary criteria are included in the selection method. For example:

a. The statement about not selecting from the authors own previous replication work is a pure redundancy from the second (more general) criteria, which stated that previously replicated papers were to be selected. By definition, if the author had previously replicated a study then someone had previously replicated it.

The timeframe and journals that I listed reflect the sampling frame that I used to identify papers that I could possibly replicate in my work, Chang and Li (2015, 2017, Forthcoming), not necessarily the actual samples of papers from that sampling frame that I replicated, so this statement is not redundant with the earlier statement about the replication wiki. For example, Gourinchas and Obstfeld (2012) was in my sampling frame (published in American Economic Journal; Macroeconomics in 2012), but I did not attempt a replication of Gourinchas and Obstfeld (2012). Of course, if all of the papers from the sampling frame in Chang and Li (2015, 2017, Forthcoming) also had replications registered on the replication wiki, which would have been conducted by other researchers, then this statement is redundant, but I did not check to see whether this was the case on March 14, 2017. I will attempt to clarify this point in a future draft.

b. To exclude work by those with a connection to the current place of employment and those with a personal correspondence history seems completely arbitrary.

When I started into the description of how the paper to target was "selected", I suspected some sort of broad/general criteria that would produce a large population of potential choices that was not subject to obvious biases, and then with that pool some sort of random selection procedure would be applied. [For example, alphabetize the entire list of articles then use a random number generator to select one by position on the list.]

c. The criteria of "A paper that I read within a year prior to the special issue's call" illustrates the arbitrary nature of this discussion even more clearly than points a and b above. I do not actually know how to interpret this. Is the author really suggesting others follow this approach? How does this 'selection criteria' not boil down in the end to picking a paper the individual author is comfortable with and interested in for independent reasons? Again, I am not challenging the idea that this 'comfortable' method is a fine way to proceed with picking a study to replicate, but my point is that any 'comfortable' and/or 'clearly non-scientific' selection method does not merit independent publication as a stand-alone academic contribution.

The special issue merely required that submitters select "*an influential economics article that has not previously been replicated,*" with a focus on applying the procedure to the candidate article (i.e., less focus on how to select an article). You're correct in that, in addition to the special issue's two mandated article selection criteria, I added a few additional, and arbitrary, criteria to select the candidate for the replication plan. I thought that the two selection criteria to mitigate bias (on disregarding papers from my own institution and papers with authors that I had corresponded with) were obvious bias reducers, but I was mistaken. I am not arguing that researchers should use these criteria to select an article to replicate, but I needed some way to pick a paper to apply a preanalysis plan to.

3. Several arbitrary and/or unnecessary criteria are included in the definition of "success". For example:

a. Point 4 is quite vague. The author would wait a "prespecified amount of time" (later quantifying as a 'few' weeks) and would engage in a "prespecified number of attempts" (never quantifying).

I will quantify the number of attempts in the subsequent draft.

b. Similarly, point 5 mentions a "flowtime of around two months" to do the replication. How is this different from a normal research not just thinking to themselves of their research goals?

I'm not sure what you mean by this question. If your question was meant to say "*How is this different from a normal researcher just thinking to themselves of their research goals?*" then the difference is that I am arguing for commitment.

Would the author's own life/circumstances that played out during the time in question not be accounted for in a reasonable way?

The preanalysis plan should attempt to account for potential roadblocks when prespecifying the amount of flowtime allocated to the replication. A researcher could prespecify that, if the researcher were to encounter some type of catastrophic circumstance (or even other circumstances that are well defined), then the flowtime allocated to the replication could be suspended.

Again, I just don't see a rigorous scientific contribution here that others could benefit from.

c. Most importantly, point 9 indicates that "If the data that I downloaded was obviously flawed, then I would give up and work on another research paper." The mind (ok, MY mind) reels upon reading this. I trust the authors of the original Haurin and Rosenthal paper because I have no reason at all not to. I trust the author of this paper for the same reason. Setting that aside, it seems that in the case of a purely doctored research endeavor, this step in the process would actually cause the research conducting a replication to give up when they should not. Assume the original study misrepresented the steps they took to include/exclude observations. When the replicator checked simple things like raw observation counts they would get something quite different and 'give up' on that replication.

I crafted the preanalysis plan with the intent of writing an extension to Haurin and Rosenthal (2007), which would take, as a necessary condition, the ability to replicate the Haurin and Rosenthal (2007) results successfully. My intent from point 9 was that only a major discrepancy in the downloaded data would cause the replication to cease, as point 8 should have checked more simple things like observation counts. Perhaps the stopping rule in point 9 would cause me to give up on the project too quickly.

You're correct that, in the case of a purely doctored research endeavor, applying this point would cause the replication to cease. As I wrote the replication plan with the intent of writing an extension, not necessarily with the goal of verification or uncovering potentially doctored research, I would not extend the argument to "should not give up".

If, instead, the intended goal of the replication was to verify Haurin and Rosenthal (2007), then point 9 may not be an appropriate stopping rule.

4. While I understand the researcher, and in turn the journal to which the paper is submitted to, carries a direct interest in the field of Economics. However, Economics has direct connections to other disciplines including Finance, Accounting, Marketing, Political Science, Geography, Sociology, Urban Planning, and many others. [A simple review of the 750+ journals indexed by Econlit substantiates this point.] In every way, the paper was written as if other fields do not exist. For some topics, I could see that an acceptable choice, *but in this case is difficult to defend given the position of the hard sciences, medicine, and psychology as being so dramatically ahead of Economics in terms of replication protocol. Since the paper does not shy away from making normative assertions, I'll make one of my own. Academic research should seek to reflect the current state of knowledge on the investigated topic, regardless of what field has produced the key insights that determine that knowledge. For example, relevant classic research from psychology that considers the key differences between replications and the scientific method in physical sciences relative to social/behavioral sciences is ignored.*

I'm not sure how to respond to this comment exactly. When I argued for using preanalysis plans that specify flowtime, budget, and intended results for "success" in replications, I was not thinking directly of economics per say. But as economics is the field that I am most familiar with, the references fall within economics, and the special issue's requirement mandated selection of an economics article to apply the proposed replication procedure to. If you have any specific suggestions on how I could/should broaden the paper, then I welcome them.

Kind Regards,

Andrew C. Chang

Senior Economist, Division of Research and Statistics

Board of Governors of the Federal Reserve System

a.christopher.chang@gmail.com

References

Chang, Andrew C., and Phillip Li. "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not"," Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System (2015).

Chang, Andrew C., and Phillip Li. "A Preanalysis Plan to Replicate Sixty Economics Papers that Worked Half of the Time," American Economic Review 107:5 (2017), 60-64.

Chang, Andrew C., and Phillip Li. "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Often Not"," Critical Finance Review (Forthcoming).

Gourinchas, Pierre-Olivier, and Maurice Obstfeld, "Stories of the Twentieth Century for the Twenty-First," American Economic Journal: Macroeconomics, 4:1 (2012), 226-265.

Haurin, Donald R., and Stuart S. Rosenthal. "The Influence of Household Formation on Homeownership Rates Across Time and Race," Real Estate Economics 35:4 (2007), 411-450.