

Chang, Andrew C.

Working Paper

A replication recipe: List your ingredients before you start cooking

Economics Discussion Papers, No. 2017-74

Provided in Cooperation with:

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Chang, Andrew C. (2017) : A replication recipe: List your ingredients before you start cooking, Economics Discussion Papers, No. 2017-74, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:

<https://hdl.handle.net/10419/169133>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

A replication recipe: list your ingredients before you start cooking

Andrew C. Chang

Abstract

The author argues that researchers should do replications using preanalysis plans. These plans should specify at least three characteristics: (1) how much flowtime the researchers will spend, (2) how much money and effort (working hours) the researchers will spend, and (3) the intended results and the precision of the replication necessary for “success”. A researcher’s replication will be “successful” according to context-specific criteria in the preanalysis plan. The author also argues that the two biggest drawbacks of preanalysis plans—(1) that they discount unexpected but extraordinary findings and (2) that they make it difficult for researchers to prespecify all possible actions in their decision trees—are less relevant for replications compared with new research. The author concludes with describing a preanalysis plan for replicating a paper on housing demand and household formation.

(Published in Special Issue [The practice of replication](#))

JEL B41 C80 C81 R21

Keywords Data and code files; household formation; housing demand; preanalysis; prespecification; publication bias; replication

Authors

Andrew C. Chang, ✉ Board of Governors of the Federal Reserve System, Washington DC, USA, a.christopher.chang@gmail.com

The views and opinions expressed here are the author's and are not necessarily those of the Board of Governors of the Federal Reserve System. The author prepared this paper for a special issue on definitions of “successful” and approaches to replications, with an example of how he would conduct a replication. The author thanks Christopher Karlsten for helpful comments. Any errors are the author's responsibility.

Citation Andrew C. Chang (2017). A replication recipe: list your ingredients before you start cooking. Economics Discussion Papers, No 2017-74, Kiel Institute for the World Economy. <http://www.economics-ejournal.org/economics/discussionpapers/2017-74>

1 Replication Recipe

Researchers should do replications using preanalysis plans that set at least three criteria: (1) the flowtime of the intended replication; (2) the budget of the replication, in terms of money and working hours; and (3) the set of estimates and the degree of precision that will define a “successful” replication.

Prespecification of the amount of flowtime and the budget of a replication anchors these dimensions of a replication. Without prespecification, the amount of flowtime and budget that you could invest in a replication could grow uncontrollably.¹

The context of the replication and your preferences will determine how much flowtime and budget you are willing to invest to get a “successful” replication. The Bill & Melinda Gates Foundation is willing to give the International Initiative for Impact Evaluation (or 3ie) a large budget to do aid impact replications both because the foundation has money and because the foundation’s aid projects depend on accurate research. Your budget for a “successful” replication is, most likely, less than that of the Bill & Melinda Gates Foundation. Context should also determine what you define as a “successful” replication, in terms of both the set of results from the original paper that you are interested in and the precision of the replication of those results.

I know that “context matters” is an unsatisfactory answer for a special journal issue seeking a definitive answer of what defines a “successful” replication. But context *does* matter.

Consider screaming, “*The roof is on **fire!***” When might the context determine the meaning of this outburst?

How about the following:

1. You are partying at a nightclub on New Year’s Eve.
2. You are on the phone with fire and rescue dispatch.

¹Prespecification of a sufficient flowtime and budget also mitigates any inclinations you might have for debunking the original paper.

3. You are a contestant on *Wheel of Fortune*.

Now consider three contexts where you might want to do a replication and where the context determines what a “successful” replication means:

1. You are verifying the original paper for the archival record.
2. You are writing a new paper that extends the original paper.
3. You are learning an econometric technique from the original paper.

How would you *verify an original paper for the archival record*? Presumably, you would want to replicate all estimates that appear in the main text to machine precision.² You would be less concerned—if you would be concerned at all—about any non-published online appendix to the paper that would not be a part of the archival record.

What about your *new paper that extends the original paper*? You probably would be interested in a subset of key results from the original paper, most likely the “main result” of the original paper. Robustness check #534 in footnote #81 would probably be irrelevant for you even if it appears in the main text of the original paper. A “successful” replication in this context would probably treat your replication of 0.41, when the original estimate was 0.43, as a “successful” replication.

And how about *learning an econometric technique from the original paper*? In this case, the technique you want to learn might be robustness check #534 scribbled out in footnote #81, making this footnote the most important part of the paper. But a “successful” replication for your human capital development is the procedure that you would go through to learn the technique, so long as you are confident that the procedure you would go through is correct. The sausage of econometric estimates might be irrelevant in this context.

²Of course, if you had a vendetta to debunk the original paper, then “success” for you would be getting estimates as far from the original paper as possible.

Much like preanalysis plans for new research, preanalysis plans for replications: (1) reduce our incentives to specification search and p-hack (that is, run models until your p-values are “significant”), (2) ground our estimates in statistical theory because the specifications that we report are not pretested (that is, the estimates are not conditioned on observing some other unreported specification), and (3) provide a results-free defense against potential criticism.³

A results-free defense is even more important for replications than for new research. At least some authors will not like you attempting to replicate their research. Should you find something contrary to the original paper, these authors may feel extra motivated to rebut your replication. Possibly with a less-than-civil rebuttal. A results-free defense will help absolve you from criticisms of replicator bias.

Preanalysis plans have two main weaknesses: (1) they force researchers to ignore findings outside the scope of the preanalysis plan, so that unexpected but extraordinary findings cannot contribute to a study, and (2) researchers have difficulty prespecifying contingencies for all possible outcomes. But these weaknesses are less concerning for replications compared with new research. With respect to ignoring findings outside the scope of the preanalysis plan, when replicating you know the general scope of results that you are aiming for prior to running your replication. Therefore, finding something extraordinary that is outside the scope of the original paper that you are replicating is less likely compared with finding something extraordinary with new research, so this feature of preanalysis plans is less of a drawback for replications.⁴ Regarding the difficulty of prespecifying contingencies, you will still need to prespecify contingencies for when you find different estimates than the original paper. But the original paper, by narrowing the research question and outlining

³The first economics research that I am aware of that used a preanalysis plan is Neumark (1999, 2001). Casey, Glennerster, and Miguel (2012) discuss the strengths and weaknesses of preanalysis plans. For a further discussion on the limitations of preanalysis plans, particularly for experiments, see Coffman and Niederle (2015).

⁴Unless you undertook a replication expecting to find something extraordinary, in which case you might have a vendetta or are extremely cynical about published research.

its methodology, sets bounds on reasonable contingencies.

2 Application to Haurin and Rosenthal (2007)

I selected Haurin and Rosenthal (2007), a paper on housing demand and household formation, as an application of how I would conduct a replication.⁵

To select Haurin and Rosenthal (2007), I used two criteria to match the special issue's requirements:

- The issue's requirement for an "influential economics article," which I took to mean a paper with at least one Google Scholar citation as of March 14, 2017.
- The issue's requirement for a paper "not previously replicated," which I took to mean a paper without a replication registered on the Replication Wiki⁶ as of March 14, 2017.

I used an additional three criteria, which the special issue did not mandate, to reduce my own biases in writing a replication plan:

- A paper not published from July 2008 to September 2013 in the following 13 well-regarded journals: *American Economic Journal: Economic Policy*; *American Economic Journal: Macroeconomics*; *American Economic Review*; *American Economic Review: Papers and Proceedings (or P&P)*; *Canadian Journal of Economics*; *Econometrica*; *Economic Journal*; *Journal of Applied Econometrics*; *Journal of Political Economy*; *Review of Economic Dynamics*; *Review of Economic Studies*; *Review of Economics and Statistics*; and *Quarterly Journal of Economics*; which are the journal issues covered in my own work on replication, Chang and Li (2015a, 2017, Forthcoming).

⁵I did not tell the editor that I had self-imposed any of the criteria in this section.

⁶See Höfler (2017), http://replication.uni-goettingen.de/wiki/index.php/Main_Page

- A paper not written by an author who, at the time of the paper’s publication, was from the Board of Governors of the Federal Reserve System.
- A paper not written by an author with whom I had corresponded before March 14, 2017.

Finally, I used two characteristics to reduce my search costs and ensure feasibility of the replication plan.

- A paper that I read within a year prior to the special issue’s call.
- An empirical paper that used no confidential or proprietary data.

3 Replication Steps and Discussion of “Success”

I focus this discussion on the context of trying to replicate the results of Haurin and Rosenthal (2007) in order to write an extension.

The steps that I would take would be as follows:

1. I would want to replicate their results using my code and their raw data,⁷ since my extension would take Haurin and Rosenthal (2007)’s results as a necessary condition. I would want their raw data, not their transformed data, so that I could code everything from start to finish.
2. As I would be planning on coding Haurin and Rosenthal (2007) but could still make a coding error in doing so, I still would want Haurin and Rosenthal (2007)’s code for verification purposes.

⁷I would want to use their data because public-use data sets are revised over time and authors rarely document which version of data they use (Chang and Li, 2015b), so I would want to eliminate data revisions as a source of uncertainty. In addition, even if you know which version of the public-use data to look for, locating a particular version of a public-use data set can be time consuming or even impossible.

3. I would search for Haurin and Rosenthal (2007)'s raw data and code in standard repositories: the journal's website, the author's website, the Inter-university Consortium for Political and Social Research (or ICPSR), etc.
4. If I could not find Haurin and Rosenthal (2007)'s code or raw data, then I would contact the authors to request their files. I would wait a prespecified amount of time for the authors to respond to my inquiries (say, a few weeks) and also set a prespecified number of attempts that I would try to contact them.
5. I would estimate a flowtime of around two months to do this replication. Some of that flowtime estimate is for requesting their files and just waiting for responses. The remainder of the flowtime is for doing actual work.
6. I would estimate about 180 hours of work (6 productive hours per working day x 30 working days).
7. For identifying the results I would want replicated, I could try replicating their Figures 1, 2, 3, 7, and 8. The first two figures would primarily be for checking to make sure the moments of the data that I would have match the moments they report. On estimation results, for extending their paper, I would be looking to see whether their basic results—their estimates of age-specific correlations between homeownership and headship and their simulations of race-specific homeownership conditional on a headship formation process of a different race—are approximately correct and whether I simulated their models accurately. Therefore, I would not need to replicate each iteration of their results for race-specific homeownership profiles. One example figure that shows that I simulated their model accurately would likely be sufficient.
8. In the scenario where I could not obtain any files that the authors used, I would follow the steps outlined in Haurin and Rosenthal (2007) to re-download their

data. I would do a cursory review of the newly downloaded data, checking to make sure the relevant variables, including those on their page 423, are in the downloaded data set, the observation counts are of similar magnitude to what they report on page 424, etc. Sadly Haurin and Rosenthal (2007) do not have a table of detailed summary statistics that I could use to compare the moments from my downloaded version of data with theirs.

9. If the data set that I downloaded was obviously flawed, then I would give up and work on another research paper. Otherwise, if the dataset passed the cursory checks, then I would work on the replication.
10. I would be “successful” if I was able to replicate the Figures that I prespecified (1, 2, 3, 7, and 8) to a reasonable degree of accuracy.

References

- Casey, Katherine, Rachel Glennerster, and Edward Miguel. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan,” *Quarterly Journal of Economics* 127:4 (2012), 1755-1812.
- 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 (2015a).
- Chang, Andrew C., and Phillip Li. “Measurement Error in Macroeconomic Data and Economics Research: Data Revisions, Gross Domestic Product, and Gross Domestic Income,” *Finance and Economics Discussion Series* 2015-102. Washington: Board of Governors of the Federal Reserve System (2015b).

- 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).
- Coffman, Lucas C., and Muriel Niederle. "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible," *Journal of Economic Perspectives* 29:3 (2015), 81-98.
- 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.
- Höfler, Jan H. "ReplicationWiki: Improving Transparency in Social Sciences Research," *D-Lib Magazine* 23:3/4 (2017). <https://doi.org/10.1045/march2017-hoeffler>
- Neumark, David. "The Employment Effects of Recent Minimum Wage Increases: Evidence from a Pre-specified Research Design," NBER Working Paper No. 7171 (1999).
- Neumark, David. "The Employment Effects of Minimum Wages: Evidence from a Prespecified Research Design," *Industrial Relations: A Journal of Economy and Society* 40:1 (2001), 121-144.

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

<http://www.economics-ejournal.org/economics/discussionpapers/2017-74>

The Editor