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 1 for "A Replication Recipe: List Your Ingredients Before You Start Cooking"

November 20th, 2017

Dear Referee 1,

Thank you for taking the time to read my paper. I wish to clarify two points:

First, I am advocating for a researcher writing preanalysis plans that contain: (1) flowtime, (2) budget, and (3) scope of work for conducting replications. These three are distinct items. The budget term, of course, includes any item with an accounting cost, but it also should include the number of *working hours* they researcher will spend. By contrast, flowtime is the total amount of time that will elapse on the project, but not actual hours spent working. For example, sending an email requesting an author's data and code files can take a replicator five minutes (budget) but waiting three weeks for a response and taking no additional actions in the interim uses up three weeks of flowtime and no additional working hours (budget).

Second, it was my intent to tie these recommendations to the "pre-commitment" literature (which, apparently, is also referred to as preanalysis, prespecification, or preregistration - but the idea being to lay out your plan of research beforehand). There are numerous advantages to prespecifying your research design, some of which are highlighted by Casey, Glennerster, and Miguel (2012), among others. Preanalysis plans limit p-hacking, limit unconscious (or even conscious) bias by the researchers, and, when applied to designing replications, they give the replicator a personal defense against accusations of bias from replicatees.[1]

Another, underappreciated, benefit of preanalysis is that reported statistical estimates are

[1]Although less widely advocated for, preanalysis plans can help limit model overfitting, which can be a consequence of p-hacking. Model overfitting has particularly disasterious consequences in forecasting (Frye, 2017).

actually drawn from the correct theoretical distributions (i.e., are not pretested). Consider the following steps a researcher might take during the course of "normal" economic analysis, where the researcher is conducting exploratory analysis that is not prespecified.

1. The researcher hypothesizes some linear relationship between two covariates $X = \{X_1, X_2\}$, and Y. The researcher then estimates the regression $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ with ordinary least squares (OLS). Lets suppose that β is identified, so $E(X'\varepsilon) = 0$.

2. After estimating $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ with OLS, the researcher finds a statistically significant coefficient on X_1 (β_1).

3. After observing a statistically significant β_1, the researcher thinks "Because I found a significant β_1, I should check to see whether a third covariate, X_3 also belongs in the model." The researcher then expands the covariate set to $X = \{X_1,\ X_2,\ X_3\}$ and proceeds to estimate $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ with OLS. Lets still suppose that, as in the case with the model with two covariates, $E(X'\varepsilon) = 0$.

4. From estimating $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ with OLS, the researcher finds significant coefficients on X_1 and X_3 and reports the OLS estimates from both the first and second regressions.

Now, by most accounts steps (1) to (4) are perfectly normal in the course of exploratory economic research. The researcher has even been perfectly forthcoming and reported both specifications.

The question is: are the estimated parameters from the second regression, $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, appropriately reported as OLS estimates? **No, they are not**.

In the steps that I've outlined, what the researcher has done is estimate a two-stage model where the second stage with $X = \{X_1,\ X_2,\ X_3\}$ would have been censored (i.e., the researcher would not have estimated the model) given a particular range of insignificant parameter estimates in the parsimonious first stage where $X = \{X_1,\ X_2\}$. **The second**

stage is pretested based on the first stage (Poirier, 1995). Yet the researcher's OLS estimates treat each model as if they were independent.

One solution to ensure that the parameter estimates from both of these regressions are appropriately reported as OLS estimates is to prespecify. If, before conducting the analysis, the researcher commits to estimating both $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ and $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, then OLS estimates are appropriate for both regressions.[2]

Let me now turn to the two largest weaknesses of preanalysis plans, both of which you mention in your report: (1) it is difficult for researchers to prespecify all necessary features of the research plan (as you correctly put it, "few research projects evolve as initially planned"), and (2) prespecification discounts extraordinary but unexpected findings (the "serendipity principle in scientific research").

On the first weakness (difficulty of prespecification), fortunately for replications the set of reasonable choices is constrained by the original article. The number of choices the would-be-replicator would have to make is still large, but not as large as for an original research paper.

On the second weakness (discounting serendipitous findings), I disagree that prespecification is antithetical to scientific research merely because serendipitous findings are discounted. Prespecification is, by design, supposed to enhance the validity of science through constraining (though not eliminating) researcher bias. Prespecification is also being, or has been, embraced by non-economics disciplines. For example, since 2007 U.S. clinical trials that meet the FDAAA 801 definition have to be prespecified (National Institutes of Health, 2017a) and research published in journals that belong to the International Committee of Medical Journal Editors is required to be prespecified (National Institutes of Health, 2017b). Prespecification is even required for clinical trials with no external sources of funding (National Institutes of Health, 2017b), so the requirement for prespecification in clinical trials is not driven merely

[2]An alternative solution to this problem is to have an estimator that models both the first stage and second stage as a censored outcome of the first stage, but even such an estimator would need to be employed prior to simply running OLS on the first regression.

by the requirements of a granter. Does the fact that the research from U.S. clinical trails are prespecified mean that U.S. clinical trials are antithetical to science? I would hope not.

Again, thank you for taking the time to review my paper.

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

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.

Frye, Jon, "The Etiology of the Overfit Model," Working Paper (2017).

National Institutes of Health. U.S. National Library of Medicine. ClinicalTrials.gov FDAAA 801 Requirements. https://clinicaltrials.gov/ct2/manage-recs/fdaaa#WhichTrialsMustBeRegistered. Accessed November 10th, 2017a.

National Institutes of Health. U.S. National Library of Medicine. ClinicalTrials.gov Frequently Asked Questions. https://clinicaltrials.gov/ct2/manage-recs/faq. Accessed November 10th, 2017b.

Poirier, Dale J. Intermediate Statistics and Econometrics: A Comparative Approach. MIT Press (1995).