Documenting Quantitative Research to Ensure Computational Reproducibility

Richard Ball
Professor of Economics, Haverford College
Director, Project TIER

47th annual meeting of the Society for Neuroscience
Washington, DC
13 November 2017

Project TIER is supported by the Alfred P. Sloan Foundation

This talk is about *computational reproducibility*:

> reproducing results reported in published research by replicating the computations by which they were originally generated

The talk is based on my experiences with computational reproducibility in economics and other social sciences, but the core principles apply to the natural sciences, including neuroscience.

A "*transparency movement*" is underway in the social sciences.

Broadly, this movement promotes ***openness*** in the conduct and communication of research.

The purposes of increasing transparency and openness are to:

- Increase the validity and credibility of research findings
- Foster collaborative and cumulative progress in social science research

Dimensions of the research transparency movement

The research transparency movement addresses many issues, notably:

- Sample size and power
- Project registration and pre-analysis plans
- Computational reproducibility

Dimensions of the research transparency movement (continued)

- Experimental replicability
- P-hacking
- Publication bias

BITSS (The Berkeley Initiative for Transparency in the Social Sciences) is a major leader and clearinghouse of information for this movement.

The BITSS website (www.bitss.org) in an excellent place to start if you would like to learn more. In particular, see

Ted Miguel's spring 2015 graduate course on research transparency—syllabus and videos of 14 lectures
http://www.bitss.org/education/economics-270d/

Miguel and Christensen, forthcoming in the *Journal of Economic Literature*
http://emiguel.econ.berkeley.edu/assets/miguel_research/78/Transparency-JEL-2016-12-20.pdf

BITSS MOOC
https://www.bitss.org/events/mooc-transparent-and-open-social-science/

Computational reproducibility is a *sine qua non* of any other dimension of research transparency.

Indeed, how can anyone not be flummoxed by the idea that computational reproducibility cannot be taken for granted in any field of research that aspires to any reasonable standard of rigor or credibility?

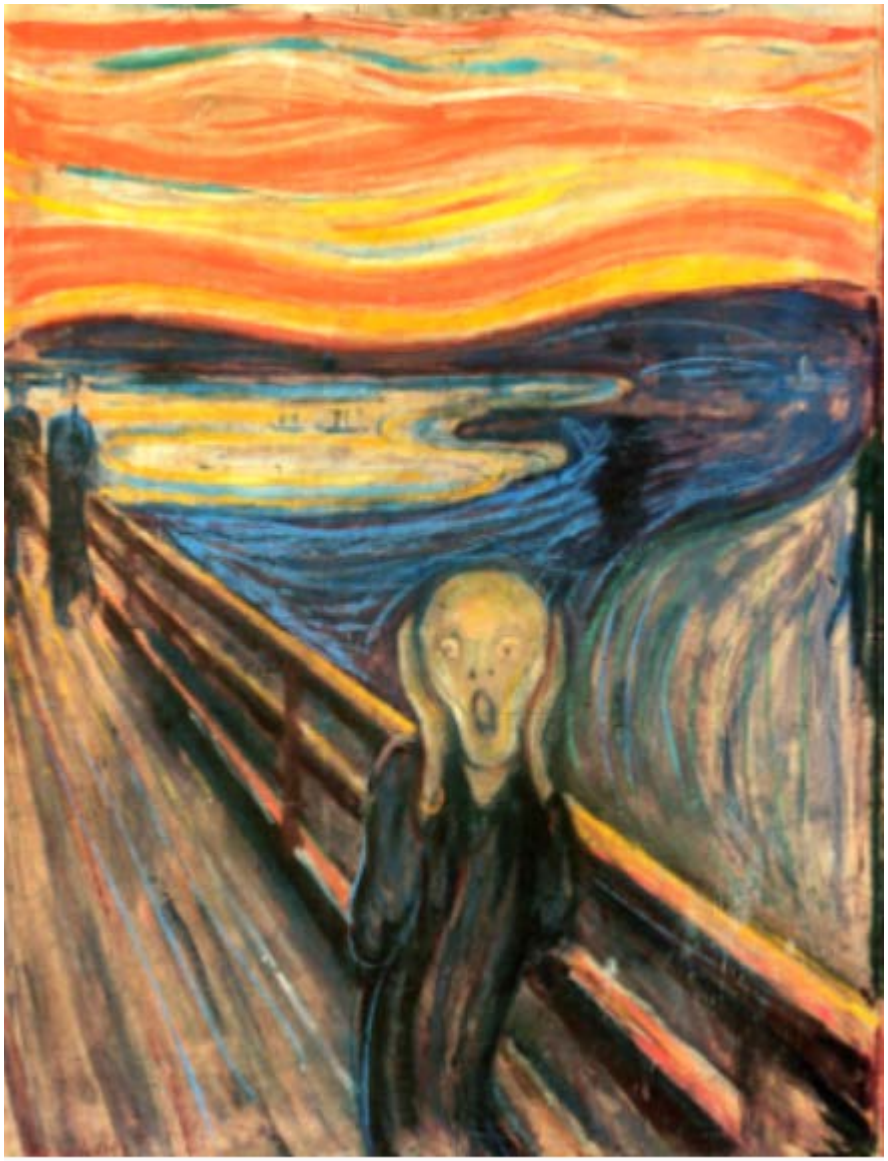
But in fact, there is a great deal of evidence that the computational results reported in a large proportion of the research conducted in many fields across the social and natural sciences cannot be reproduced.

Evidence on the reproducibility of empirical research in economics

The Big Bang: The 1986 *Journal of Money, Credit and Banking* (*JMCB*) Project

> Dewald, W.G., Thursby, J.G., & Anderson, R.G. (1986). Replication in Empirical Economics: The Journal of Money, Credit and Banking project. *The American Economic Review*, 76(4), 587-603.

- 154 requests for data and code
- Received 90 responses with some information
- Reviewed the first 54 cases
- Only 8 reproduced the results reported in the paper (15%)



Numerous follow-up studies show the problem persists

A few examples:

McCullough, Bruce D., Kerry Anne McGeary, and Teresa D. Harrison (2006). "Lessons from the *JMCB* Archive," *Journal of Money, Credit and Banking* 38(4): 1093-1107.

McCullough, Bruce D., Kerry Anne McGeary, and Teresa D. Harrison (2008). "Do Economics Journal Archives Promote Replicable Research?" *Canadian Journal of Economics* 41(4): 1406-1420.

Hoeffler, Jan (2014). "Teaching Replication in Quantitative Empirical Economics." Presented at the Meetings of the European Economic Association and the Econometric Society, Toulouse, France, August 28. http://www.eea-esem.com/eea-esem/2014/prog/viewpaper.asp?pid=3108.

Chang, Andrew C., and Phillip Li (2015). "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, http://dx.doi.org/10.17016/FEDS.2015.083.

Despite the initial anemic response to the findings of the *JMCB* Project, there has been some gradual movement toward better practices

- Many journals in economics and other social sciences require authors to submit replication data and code

  See, for example, the current "data availability policy" adopted by all the research journals of the American Economic Association. https://www.aeaweb.org/journals/policies/data-availability-policy

- These policies are on the right track: ***documentation is the key to computational reproducibility***

An instructive example from neuroscience:

Author-requested retraction of Anderson *et al.* (2014). "Induced Alpha Rhythms…". *Journal of Neuroscience* 4(22):7587–7599.

All analyses were performed using MATLAB in conjunction with the Signal Processing and EEGLAB Toolboxes. The raw EEG signal was bandpass filtered into a range of frequency bands using a two-way least-squares finite impulse response filter (eegfilt.m from EEGLAB toolbox; Delorme and Makeig, 2004). This filtering method uses a zero-phase forward and reverse operation, which ensures that phase values are not distorted, as can occur with forward-only filtering methods. A Hilbert transform (MATLAB Signal Processing Toolbox) was then applied to the frequency-specific filtered waveforms to extract instantaneous power values. The Hilbert Transform produces the complex analytic signal, $z(t)$, of the filtered EEG, $f(t)$, where $z(t) = f(t) + i\tilde{f}(t) = A(t)e^{i\varphi(t)}$, from which instantaneous amplitude, $A(t)$, was extracted; $\tilde{f}(t)$ is the Hilbert Transform of $f(t)$ and $i = \sqrt{-1}$. Power was estimated at each time point using the following MATLAB syntax:

```
abs(hilbert(eegfilt(data,Fs,f1,f2))).^2,
```

where data is a 2D matrix of raw EEG (# of trials × # of samples), Fs is the sampling frequency (250 Hz), f1 is the lower bound of the filtered frequency band, and f2 is the upper bound of the filtered frequency band.

The retraction:

The Journal of Neuroscience, February 11, 2015 • 35(6):2838

Author-Initiated Retraction: Anderson et al, Induced Alpha Rhythms Track the Content and Quality of Visual Working Memory Representations with High Temporal Precision

At the request of the authors, *The Journal of Neuroscience* is retracting "Induced Alpha Rhythms Track the Content and Quality of Visual Working Memory Representations with High Temporal Precision" by David E. Anderson, John T. Serences, Edward K. Vogel, and Edward Awh, which appeared on pages 7587–7599 of the May 28, 2014 issue.

We regret that there was an error in the analytic code used to compute oscillatory power in our article. Specifically, there was a matrix transposition error in the code (see abs(hilbert(eegfilt(data,Fs,f1,f2))).$\hat{2}$ on page 7588, right column, end of second full paragraph). The data matrix was oriented correctly for the call to eegfilt, but the output of the call to eegfilt was not correctly transposed in the standard Matlab format before passing into the built-in Matlab 'hilbert' function, as the EEGLAB function 'eegfilt' and the built-in Matlab function 'hilbert' require the data matrix to have different dimensions in order to operate correctly across time. Fortunately, this error

Documenting Computational Research

For references focused on neuroscience, see (among others):

Gilmore, R. *et al.* Progress Toward Openness, Transparency, and Reproducibility in Cognitive Neuroscience. *Ann N Y Acad Sci*. 2017 May ; 1396(1): 5–18. doi:10.1111/nyas.13325.

Eglen, S. *et al.* Towards standard practices for sharing computer code and programs in neuroscience. bioRχiv preprint, February 28, 2017. https://doi.org/10.1101/045104.

References focused on neuroscience (continued)

Gorgolewski, K and Poldrack R. A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research. *Plos Biology*. 2016 Jul 7;14(7):e1002506. doi: 10.1371/journal.pbio.1002506.

Stanford Center for Reproducible Neuroscience

http://reproducibility.stanford.edu/

Principles and Practices from the social sciences

Highlights, from big concepts to nitty-gritty

Purposes

Replication documentation that accompanies an empirical research paper is valuable to the extent it facilitates:

(i) Confirmation

(ii) Robustness checking

(iii) Extension

(iv) Communication

Principles

(i) Complete replicability

(ii) Independent replicability

(iii) Automated and portable replicability

(iii) Realism and empathy

The need for research compendia

From http://ropensci.github.io/reproducibility-guide/sections/introduction/:

What are the principles of reproducible research?

The motivating principle for reproducible research is that the traditional unit of scholarly communication - a published article - is only the tip of the iceberg of the research process. Jon Claerbout described the article as merely an advertisement for research (Claerbout and Karrenbach 1992):

> An article about computational results is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.

Keys to fully automated, portable replication documentation:

- Reify the folder structure. The structure of folders and subfolders in which you store the documentation, and where in that directory structure each of your files is stored, is itself an integral part of the documentation.

- Be explicit about the working directory. For each command file, designate one of the folders within your documentation as the one that should be set as the working directory when the command file is executed.

- Use relative directory paths. In your command files, indicate the folders in which files that need to be accessed are located or files that need to be saved should be stored with directory paths specified relative to the working directory.

For a complete set of guidelines, see the DRESS Protocol (Documenting Research in the Empirical Social Sciences)

http://www.projecttier.org/tier-protocol/dress-protocol/

An example:

Documentation of Price, J. and Wolfers, J. "Racial Discrimination Among NBA Referees." Quarterly Journal of Economics, 125(4):1859-1887

https://osf.io/v5tn7/

What about confidential or proprietary data?

Restricted data pose a challenge to ***data citation***, not to ***research documentation*** and ***reproducibility***.

The FAIR principles address this issue; also ongoing work at the Harvard Institute for Quantitative Social Sciences.

The win-win nature of conducting research transparently and reproducibly

The steps you need to take to ensure the reproducibility of your research also benefit you while you are working on a project.

- Communicating with collaborators and/or advisors.
- Communicating with your future self.
- Establishing an efficient and flexible workflow: you will never lose track of your stuff or have trouble remembering where you left off the last time you worked on the project.
- Revise and re-submits.
- Easily answer questions posed about your research after publication—or never have to field them at all.

Project TIER is focused on educating the next generation of scholars.

www.projecttier.org