A Standard for Replication in Computational Social Science

Problem statement

Public policy depends on sound social science (particularly, but not exclusively, economics). In the past couple of decades, economists have come to rely heavily on results derived from large and difficult computational modeling - a trend that accelerated sharply after the failure of simpler traditional models to explain much about the Great Financial Crisis of 2008-11. But such modeling is largely accomplished using code written by the economists themselves (or research assistants), who have little incentive (and perhaps little capacity) to produce code that is runnable by anyone else (or even by themselves a few years down the road).

So far, only a few published papers in top journals have been shown to have explicit computational errors that affected the economic conclusions, but this is probably because not many replication attempts have been made. It is so difficult to replicate and then build on someone else's work that scholars mostly start from scratch (or from their own private codebase accumulated over decades of experience). The field is thus in serious danger of experiencing a "replicability crisis" of the kind that has devastated the credibility of many other fields (especially in the social sciences).

Solution

We aim to develop and promote an open electronic publication format for computational social science research that will guarantee reproducibility and robustness, and will be easy to use. This format will enable: simple, turn-key execution (on *any* computer with sufficient resources) of the code that produces the paper's claimed results; benchmarking, and robustness tools; and standardized metadata to document the project's purpose, provenance, and conclusions.

This project has grown out of the REMARK fork of the Econ-ARK project. Originally, REMARKs were developed as a way to share results obtained using Econ-ARK's other fork, the HARK toolkit. But after discussions with "data editors" of top journals, we have realized that the existing REMARK format could evolve into a common standard for journal submissions, and for their replication archives. Our aims dovetail with those of many Sloan-funded projects: Open Journals, Open Source Program Offices, data citation standards, scholarly communication, and better software for science, and others.

We have been in communication with a team of people who intend to establish a *Journal of Open Source Economics*, ('JOSEcon') and have been developing the REMARK standard with the goal of meeting the requirements that the editors of such a journal would need for a submission. A mockup of the result of our collaboration with JOSEcon is now available live and online.

Challenges

- *Commensurability of scientific results*. Unlike other fields, the computational social sciences have not standardized a way to compare substantive results across multiple implementations. While there is some precedent for replication competitions, there is no common infrastructure to support these.

- *Accommodating multiple programming language ecosystems.* While there are many *de facto* standards and conventions for reliable software and computation, these have mostly been developed in industry, government (NASA; NIST), and university libraries and have been little used by social scientists. The result is a Tower of Babel, which complicates every part of the research process (and reviewing and editing).

- *Standardization and adoption.* For a format for reproducible research to become a standard, it must be accepted by a standards development organization after getting buy-in from multiple stakeholders. Ultimately, our goal is for the format, or standard, to be adopted by major academic journals in a variety of disciplines.

Activities

- *Establish a format for distinguishing which results should be reproducible*. As part of our proposed format, we will encourage authors to identify which results of their computation should be replicable by others using the same modeling framework.

- *Define a format for benchmarking and scientific testing.* Such tools enable researchers to work cumulatively towards improved computational methods on known problems.

- *Hosting of replication competitions*. We will run competitions using the format, inviting participants to replicate the results of some well-known papers. We anticipate the results will be publishable in journals, and be a way of spreading the format as a standard.

- *Educational tutorials.* To promote the format, we will run tutorials at several major conferences.

- *Formation of an independent board of advisors*. We will recruit and convene an independent board of advisors composed of computational scientists, journal data editors, and library scientists. The project will be independent of Econ-ARK.

Budget

$450,000 over three years.