

PEDAGOGY FOR DOCUMENTATION AND REPLICABILITY OF
EMPIRICAL RESEARCH: EXPERIENCES AND AN AGENDA

or

PROJECT TIER: WHENCE? WHY? AND WHITHER?

Richard Ball
Associate Professor of Economics
Haverford College

Norm Medeiros
Associate Librarian
Haverford College

www.haverford.edu/TIER

Presentation at the NSF Workshop
on the Robustness of Social Science Research
February 20-21, 2014

WHENCE?

Project TIER developed organically in the context of an introductory statistics class for undergraduate economics majors.

Experience gained over several years of assigning, advising and reading student research papers for the class led to the development of a set of instructions—a pdf document—for assembling comprehensive documentation for a research paper.

We call these instructions a "protocol."

The protocol specifies a set of electronic documents that students submit at the same time they turn in their completed research papers.

The files specified by the protocol include:

all original data files from which data for the project were extracted

metadata required to understand and interpret the contents of the original data files (e.g., pdfs of, or links to, codebooks or users' guides)

command files containing code that executes all the steps required to get from the original data files to the empirical results reported in the paper, including

importing original data files; cleaning, merging, and all other processing required to construct the "final data set(s)" used for analysis; generating all the statistical results, tables and figures presented in the paper

a readme file that describes all the documents included in the electronic documentation

The principle underlying the protocol:

Documentation of an empirical research project should allow

easy

exact

soup-to-nuts

replication of all reported statistical results.

Our students now routinely produce documentation that meets this standard, both for papers for introductory statistics classes and for senior theses.

WHY?

There are pedagogical benefits to teaching students to document their research that are independent of the broad issue of robustness of social science research:

Students organize their work much better and understand what they are doing much more clearly when they have to do everything in command files.

I can advise them much more effectively throughout the project when I can get my hands on their data and code and play around with them. Similarly, I can evaluate and comment on their final papers much more insightfully.

When students keep track of and understand what they are doing with their data, they gain confidence that they can learn and discover real, meaningful things.

The entire documentation exercise reinforces some fundamental educational and life principles: transparency; integrity; understanding how you arrive at the claims you make and knowing how to justify or substantiate them.

Introducing students to good practices for documentation and replicability can also contribute to the enhancement of the robustness of professional social science research.

Most obviously, tomorrow's researchers will come from the ranks of today's students.

But there are other, somewhat nearer-term, ways that this initiative can have some positive spill-overs to professional research practice. I will mention some of these in a moment.

We are now engaged in a variety of efforts to disseminate what we have learned through Project TIER to instructors at other institutions.

With the support of an ICPSR/Sloan Challenge Grant for academic year 2013-14, we are:

hosting two workshops for nationally-recruited faculty to demonstrate the way we introduce students to our documentation protocol, and to encourage them to consider steps they could take to introduce their students to good documentation practices

speaking at conferences, writing articles

developing our web page

developing a strategy for continuing these efforts after this year

WHITHER?

Our agenda consists of three strategic components:

Infrastructure development

Dissemination and outreach

Developing partnerships with other efforts to promote the many dimensions of robustness

Infrastructure:

development of the protocol: generalization and refinement of the instructions we give students for documenting their research projects

integrating our website with a platform (Dataverse?) for managing and sharing data and documentation

Dissemination:

continued outreach: conferences, workshops, webinars

incentives and support for faculty to teach documentation:

an annual meeting to share experiences

recognition or symbolic status ("Professor Smith is an XYZ Foundation Fellow for the Promotion of Responsible Research Practices")

stipends for faculty introducing instruction on documentation into their classes and research advising

Partnerships:

The kind of replication Project TIER is focused on is sometimes called "computational reproducibility."

There are other kinds of replication, and many other dimensions of the broad notion of "robustness."

The various initiatives that concentrate on different dimensions of robustness are highly complementary. They are motivated by common underlying values.

Are there ways of fostering collaboration among groups working on different dimensions of robustness to leverage these complementarities?

SPILL-OVERS AND IMPLICATIONS FOR PROFESSIONAL PRACTICE

The demonstration effect:

If college sophomores are constructing comprehensive, soup-to-nuts documentation for their intro stats class research projects, how hard can it be?

The value of simply writing down a clear and sufficiently elaborated protocol:

If you build it, they (or at least some of them) will come.