

www.haverford.edu/tier

TEACHING INTEGRITY IN EMPIRICAL RESEARCH:
A SOUP-TO-NUTS PROTOCOL FOR DOCUMENTING
DATA MANAGEMENT AND ANALYSIS

Richard Ball
Associate Professor of Economics
Haverford College

Norm Medeiros
Associate Librarian
Haverford College

www.haverford.edu/tier

Presented at the 2013 Stata Conference

New Orleans, July 18, 2013

A SCARY FACT

Most empirical research published in professional economics journals is not adequately documented.

And the statistical results reported in most professional economics journals cannot be replicated:

- not by a third party

- often not even by the author(s) of the paper

This scary fact is well-documented. For some examples, see

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

McCullough, B., McGeary, K. and Harrison, T. (2006). Lessons from the JMCB Archive. *Journal of Money, Credit and Banking* 38(4):1093-1107.

McCullough, B. and McKittrick, R. (2009). Check the Numbers: The Case for Due Diligence in Policy Formation. Fraser Institute Studies in Risk and Regulation.

Glandon, P. (2012). Report on the *American Economic Review* Data Availability Compliance Project. *American Economic Review* 101(3): 695-699.

Some journals now have “data policies” and maintain on-line “data archives.”

But these generally do not work well:

There is usually little enforcement or quality-control.

The standard the documentation must meet is very low:

Only the “final data set” and code that uses it to produce reported results must be submitted.

Complete documentation of original data sources and construction of final data sets not required.

For example, see the data policy for the *American Economic Review* at www.aeaweb.org/aer/data.php.

Professional economists (including journal editors) tend to resist attempts to change norms and practices concerning documentation of empirical research.

But...

...our students have to do whatever we tell them to.

We have developed a protocol for comprehensively documenting all steps of data management and analysis conducted for an empirical research paper.

We have been teaching it to our students for about seven years, and now it is used routinely

by students in our introductory statistics classes, for research papers they write for the class

and by economics majors writing empirical senior theses.

The protocol specifies a set of electronic documents that students assemble as they conduct their research, and submit when they turn in their final papers or theses.

Details of the protocol can be found in the “Student Instructions” posted at www.haverford.edu/tier.

More details and discussion can be found in a paper we have written:

Ball, R. and Medeiros, N. (2012). Teaching Integrity in Empirical Research: A Protocol for Data Management and Analysis. *Journal of Economic Education* 43(2): 182-189.

Some notable features of the protocol:

The documentation should contain everything necessary to allow a third party to reproduce all the statistical results reported in the paper, easily and exactly.

The standard of documentation is “soup-to-nuts”: from data files in the format in which they were originally obtained to the generation of the results reported in the paper, with all intermediate steps included.

The protocol is a work in progress; we are continually refining it. There are many ways it could still be extended and improved—and we would love your help!

Teaching this protocol to students has two kinds of benefits:

Enhancing professional norms and practices through a “trickle-up” effect

And there are additional pedagogical benefits. When students are required to document their work:

They understand much better what they are doing.

Instructors are able to give much better guidance and feedback.

Principles of integrity, transparency and accountability are reinforced.

For both these reasons, we have embarked upon an outreach effort that we call Project Tier (Teaching Integrity in Empirical Research).

The main goals of Project Tier:

Extend and refine the documentation protocol we have developed.

Publicize the protocol to statistics instructors at other institutions, and encourage them to teach it to their own students.

Main goals of Project Tier (continued):

Foster the development of a network of instructors committed to promoting the widespread incorporation of methods for responsibly documenting empirical research into the curricula of the social sciences.

For academic year 2013-14, Project Tier has been awarded a small grant from ICPSR and the Sloan Foundation. The grant will support:

Two faculty workshops

Talks (like this one) at conferences and professional meetings

Planning for beyond this academic year

Faculty workshops: For social science faculty who teach statistics or advise empirical student research projects

Both to take place at Haverford College

The first will be October 18-19, 2013

The second will be in the spring of 2014

We will show participants how the protocol works and discuss potentials and obstacles for others to integrate similar instruction in documentation of empirical research in their teaching and research advising.

A detailed announcement of the October workshop, and the on-line application, can be found at www.haverford.edu/tier.

We hope the workshop participants will form the core of a network of instructors committed to developing and disseminating the protocol.

Please help us recruit a motivated cohort of participants:

Recommend individual faculty members you think might be good participants.

Recommend particular institutions with programs that might be fertile ground for this project.

Let us know of outlets we should be advertising this in: list-serves, web sites, etc. (for all social sciences).

Planning for beyond this academic year

We hope that the workshops and other activities we have planned for 2013-14 will be the beginning of a longer-term effort to pursue the goals of Project Tier.

So part of our work this year is to figure out how to proceed after this year—more workshops? Talks? Webinars and on-line training for faculty?

This may mean preparing one or more proposals for further grant funding.

We would like to see a day when teaching students to assemble soup-to-nuts documentation of empirical research becomes such an integral part of the curricula of the social sciences that no one has to think about it much any more—like putting a reference list at the end of a paper.

We have a long way to go to get there, but there are signs that the glacier is beginning to slip.

The question is: What will be the most effective ways of moving us toward that day?