

TEACHING GOOD RESEARCH PRACTICE: DOCUMENTATION AND REPLICABILITY

Richard Ball
Associate Professor of Economics
Haverford College

Norm Medeiros
Associate Librarian
Haverford College

www.haverford.edu/tier

Presentation at the meetings of the
Southern Political Science Association
New Orleans, January 10, 2014

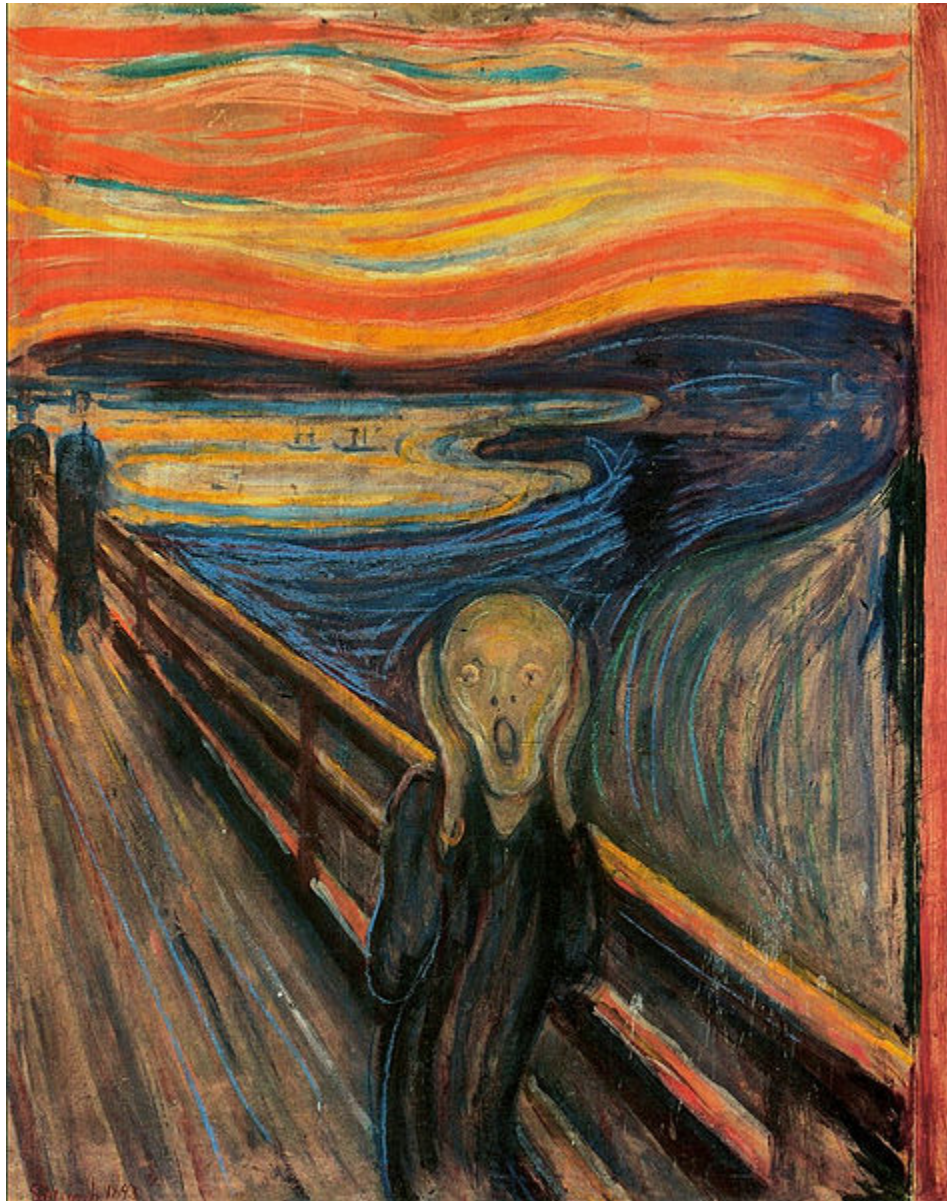
A SCARY FACT

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

In fact, the statistical results reported in most articles that appear in 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. 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.

Citations of many other studies that show how poorly documented much economic research is can be found at www.haverford.edu/tier.

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.

What to do about this state of affairs?

Convincing professional economists (including journal editors) norms, policies and practices concerning documentation of empirical research should change looks to us like a tough nut to crack.

But...

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

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

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.

We have been teaching this protocol 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.

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!

The tangible embodiment of the protocol can be found at

www.haverford.edu/TIER/protocol/

We can take a quick gander at that site...

...but an illustration is probably a better way to convey the idea.

So let's look at a student project from the spring of 2013.

Teaching this protocol to students has several 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.

Documenting your work responsibly *facilitates* your research—it is not an added burden!

For all 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.

Faculty workshops—March 2014

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

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 March workshop, and the on-line application, can be found at www.haverford.edu/tier.

The ultimate goal we are working toward:

We would like to see a *happy 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.

