Project TIER is supported by the Alfred P. Sloan Foundation.

WORKSHOP ANNOUNCEMENT

Transparent Empirical Research:
How to Make Replication Documentation Useful to Yourself and Others

Part I
Thursday September 14th, 2017
Early session: 12:00 PM – 2:00 PM
Late session: 5:00 PM – 7:00 PM
Warren 175

Part II
Friday September 15th, 2017
Early session: 11:00 AM – 2:00 PM
Late session: 2:30 PM – 5:30 PM
Warren 401

Workshop Description

This workshop will be about conducting and documenting empirical social science research in ways that ensure transparency and reproducibility. It will focus on two broad approaches to carrying out a research project: (i) copy-and-paste and (ii) dynamic documents created with a markup program. Both approaches are effective means of ensuring reproducibility, and which one a researcher chooses typically depends on her/his preferences and experience. We will highlight the win-win nature shared by the two approaches: they increase the scientific credibility of your work and its usefulness to other researchers, and simultaneously increase your productivity and reduce some of the stresses you may encounter as you conduct a research project.

The copy-and-paste approach is the most familiar to economists. The researcher writes command files for some kind of statistical package (e.g., Stata, SPSS, R, SAS, Matlab, etc.) that process the data as necessary and then generate the figures, tables, and any other results to be included in a paper. After the command files are run, it is a separate step to insert the figures and tables into the word processing (e.g, MS Word) or typesetting program (e.g., LaTex) being used to compose the paper.

Creating dynamic documents with a markup program is an alternative approach that is gaining popularity. Using a markup language (e.g., R Markdown), the researcher composes a single

document that includes both the text of the paper and code that generates the statistical results, along with tags that define the formatting of the text, figures, and tables. This markup document is then sent to an engine that renders it into a formatted paper containing both the text and results.

The sessions on Thursday, September 14 will include demonstrations of both cut-and-paste and dynamic documents approaches. We will discuss the rationale and principles underlying each approach, provide examples of projects that have been documented with each approach, and show practical tips. The copy-and-paste examples will be done with Stata and MS Word, and with Stata and LaTex; the dynamic document examples will use R Markdown. Time permitting, we will also introduce participants to two on-line platforms for managing and archiving research data, code and related documents: Open Science Framework (OSF) and GitHub. The morning and afternoon sessions will cover the same material, so participants should come to either the morning or the afternoon session.

The sessions on Friday, September 15 will be hands-on labs in which participants have the opportunity to explore whichever of the two approaches they prefer, using real data. Individuals who are engaged in or about to begin working on projects of their own are encouraged to bring whatever data, command files and associated documents they have assembled so far, and we will help them figure out how to get everything organized so that they can begin following the approaches we recommend. For individuals who don't have a current project to work on, we will provide exercises to give them some practical experience implementing these approaches to reproducible research.

At each of the Friday sessions, we will begin with some general introductory comments. After those comments, the Friday sessions will be a time for informal and independent work. Participants should feel free to come and go as they please over the course of the two sessions: work for a while on your project or exercise, take a break, come back later to work some more. Participants in the morning session are welcome to come back and continue work during the afternoon session.   Instructors will circulate to provide help with whatever you are working on.

Preparing for the Workshop
Fear not: little or no preparation is required.

No special preparation is needed for the presentations on Thursday.

If you plan to attend any of the lab sessions on Friday, it would be worthwhile to think about whether you have a project of your own that you would like to work on, and if so to begin collecting the data, code, and supporting documents that you are using so that you are prepared to work with them during the lab. More information on exactly what materials would be useful

2

to bring on Friday will be given during the presentation on Thursday. (Remember, if you do not have a current project to work on during the lab, there will be exercises available to work on.)

The following articles would make useful background reading, but they are not required. These articles are available to view or download at https://osf.io/c7mpe/?view_only=c4e9743ca7474a3ba10cef65deab7870.

Ball, R. and N. Medeiros (2012). Teaching integrity in empirical research: A protocol for documenting data management and analysis. *Journal of Economic Education* 43(2): 182-189.

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

Herndon, T., M. Ash, M. and R. Pollin (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics* 38(2):257-79.

McCullough, B.D., and R. McKitrick (2009). Check the numbers: The case for due diligence in policy formation (Fraser Institute Studies in Risk & Regulation). Retrieved from http://www.pages.drexel.edu/~bdm25/DueDiligence.pdf.

Workshop Instructors
The talks on Thursday will be presented by Richard Ball, Professor of Economics at Haverford College and Director of Project TIER, and Tomas Dvorak, Professor of Economics at Union College and a 2015-16 TIER Faculty Fellow.

The lab instructors on Friday will be Richard Ball, Tomas Dvorak, and Florio Arguillas, Research Associate at the Cornell Institute of Social and Economic Research and a 2017-18 TIER Faculty Fellow.