Tutorial 4

INF312: Worlds Become Data

Luca Carnegie

January 31, 2024

Caro (2019) writes at least 1,000 words almost every day. In this tutorial you will write every day for a week. Please pick one of the papers specified in the prerequisites and complete the following tasks:

My Paper: Koenker, Roger, and Achim Zeileis. 2009. "On Reproducible Econometric Research." Journal of Applied Econometrics 24 (5): 833–47. https://doi.org/10.1002/jae.1083.

Day 1: Transcribe, by writing each word yourself, the entire introduction.

The renowned dispute between Gauss and Legendre over priority for the invention of the method of least squares might have been resolved by Stigler (1981). A calculation of the earth's ellipticity reported by Gauss in 1799 alluded to the use of *meine Methode*; had Stigler been able to show that Gauss's estimate was consistent with the least squares solution using the four observations available to Gauss, his claim that he had been using the method since 1795 would have been strongly vindicated. Unfortunately, no such computation could be reproduced leaving the dispute in that limbo all too familiar in the computational sciences.

The question that we would like to address here is this: 200 years after Gauss, can we do better? What can be done to improve our ability to reproduce computational results in econometrics and related fields? Our main contention is that recent software developments, notably in the open-source community, make it much easier to achieve and distribute reproducible research.

What do we mean by reproducible research? Buckheit and Donoho (1995) have defined what de Leeuw (2001) has called Claerbout's Principle: "An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures." We view this as a desirable objective for econometric research. See Schwab et al. (2000) for further elaboration of this viewpoint.

The transition of econometrics from a handicraft industry (Wilson, 1973, Goldberger, 2004) to the modern sweatshop of globally interconnected computers has been a boon to productivity

and innovation, but sometimes seems to be a curse. Who among us expected to be in the "software development" business? And yet many of us find ourselves precisely in this position, and those who are not, probably should be. As we will argue below, software development is no longer something that should be left to specialized commercial developers, but instead should be an integral part of the artisanal econometric research process. Effective communication of research depends crucially on documentation and distribution of related software and data.

Some journals, such as the Journal of Applied Econometrics (JAE), support authors in this task by providing data archives (MacKinnon, 2007). However, a more integrated approach encompassing data, software, empirical analysis, and documentation is often desirable to facilitate replication of results.

Day 2: Rewrite the introduction so that it is five lines (or 10 per cent, whichever is less) shorter.

The renowned dispute between Gauss and Legendre over priority for the invention of the method of least squares might have been resolved by Stigler (1981). A calculation of the earth's ellipticity reported by Gauss in 1799 alluded to the use of *meine Methode*; had Stigler been able to show that Gauss's estimate was consistent with the least squares solution using the four observations available to Gauss, his claim that he had been using the method since 1795 would have been strongly vindicated. Unfortunately, no such computation could be reproduced leaving the dispute in that limbo all too familiar in the computational sciences.

The question that we would like to address here is this: 200 years after Gauss, can we do better? Our main contention is that recent open-source software developments make it much easier to achieve and distribute reproducible research.

But what is reproducible research in econometrics, then? Buckheit and Donoho (1995) say that reproducibility includes "…the complete software development environment and the complete set of instructions which generated the figures." We view this as a desirable objective for econometric research.

The transition of econometrics from a handicraft industry (Wilson, 1973, Goldberger, 2004) to the network of globally interconnected computers has been a boon to productivity and innovation. Most crucially, who among us expected to be in the "software development" business? However, many of us find ourselves precisely in this position, and those who are not, probably should be.

Software development is no longer something that should be left to specialized commercial developers, though. Instead, it should be an integral part of the artisanal econometric research process. To this end, effective communication of research depends crucially on documentation and distribution of related software and data. To accomplish this, a more integrated approach encompassing data, software, empirical analysis, and documentation is often desirable to facilitate replication of results.

Day 3: Transcribe, by writing each word yourself, the abstract.

Recent software developments are reviewed from the vantage point of reproducible econometric research. We argue that the emergence of new tools, particularly in the open-source community, have greatly eased the burden of documenting and archiving both empirical and simulation work in econometrics. Some of these tools are highlighted in the discussion of two small replication exercises.

Day 4: Rewrite a new, four-sentence, abstract for the paper.

New software is reviewed based on reproducible econometric research. The new software makes it much easier to document and store econometrics work. Some of this software comes from open-source creators. The software is demonstrated by replicating a couple small, older econometrics studies.

Day 5: Write a second version of your new abstract using only the 1,000 most popular words in the English language as defined by xkcd's simple writer.

The discovering people interested in people and money found stuff on the computer that helps them learn better about how people choose what to do in their life. The computer things make it easier for others like the discovering people to check the steps in their thinking. They show this by using the computer stuff to do old thinking again, checking if the answers they came up then were still right. All in all, they showed that the new things that others have made on the computer make their life easier, since it helps them create and check their work faster than before.

Day 6: Detail three points about the way the paper is written that you like

- I like the way that they use older studies initially done with less powerful mathematical tools to show that modern computers would have made their analysis a lot easier to conduct.
- In classic economist fashion, mentioning incentives as the main barrier to the adoption of these tools was an important thing to mention in the conclusion.
- Their use of headings to organize the paper by the various topics discussed was a very good idea - it made the paper a lot easier to read and digest.

Day 7: Detail one point about the way the paper is written that you do not like.

- It's written and published in LaTeX. Personally, I think LaTeX for purely textual tasks is cumbersome to use and a WYSIWYG editor is faster and more convenient.