



Data Sharing in Economics and the Role of the American Economic Review


Robert A. Moffitt


Johns Hopkins University


Editor, American Economic
Review


Data and Data Sharing in Economic Research

- While some economists are theoreticians and do not use data, the majority of economists do
- Wide variety of types of data
- (1) aggregate data, (2) individual-level administrative data (3) household surveys, (4) firm-level data, (5) experimental data


- 
- (1) aggregates (unemployment, output by industry, etc.): generally government-produced and government-collected, almost always publicly available
 - (2) individual-level administrative data (births, deaths, etc.): usually government-collected, usually publicly available though not always completely so


- 
- (3) household survey data: information on demographic and economic variables (income, employment, etc.) of individuals and households, sometimes government-collected and sometimes privately-collected, usually publicly available (without individual identifiers) but occasionally not; when privately-collected, are usually collected by large research or survey organizations (e.g., University of Michigan Institute for Survey Research), not by individual economists; individual economists almost never collect their own household survey data


- 
- (4) information on individual firms (profits, sales, revenue, etc.): sometimes government-collected and sometimes privately-collected, sometimes by individual researchers; never publicly available
 - (5) data from small-group laboratory experiments where individuals are given economic choices and their behavior is observed: often publicly available (without individual identifiers)


- 
- Data that are not publicly available are generally termed “proprietary” data
 - In the vast majority of cases, the data can be obtained by licensing and confidentiality agreements, since data are owned by organizations, not individual researchers
 - Data usually have individual or firm identifiers removed, and precautions are taken to insure that individual households and firms cannot be identified


- 
- There is a slight trend in economics toward using proprietary data to a greater degree than in the past
 - Is a result of the scarcity of large publicly-available data sets and limitations of social science funding for new ones
 - Forces a hunt for new data


- 
- New developments for proprietary data sets:
 - (1) Development of confidentiality agreements that researchers and institutions must sign agreeing not to release data to others
 - (2) Development of “enclaves”: a facility somewhere where the confidential data are put on a non-networked computer in a secure room where the researcher can work; and is not allowed to take confidential data out of the room


- 
- Even when data are publicly available, however, economists manipulate the data (subsetting, variable creation, etc.) in the process of conducting research
 - The actual analysis files, or the programs that created those analysis files from the master file, are needed to replicate someone else's research
 - Most of the focus of data sharing discussions in economics has been on this type of sharing


- 
- Consequently, data-sharing per se has not been a major issue
 - Arises per se only in the relatively small number of cases where individual economists have collected their own data and have ownership rights on it
 - The exception rather than the rule

- 
- Focus has been, instead, on research replicability
 - Different researchers often obtain different results even when analyzing the same master file
 - Different researchers often obtain different results even analyzing different data sets, for reasons not obvious
 - Some researcher's results are in actual error, and others are very sensitive to methods used, i.e., non-robust

- 
- Focus on replicability and robustness-testing has indirectly led to demand for data sharing
 - You can't replicate someone else's results or test their robustness without their data (usually; that is, unless you can “guess right” and do the same subsetting and variable creation that they did)

- 
- Historically, economists have not been very cooperative in this type of data-sharing
 - They tend not to save their programs and data from old projects, either inadvertently or intentionally
 - They tend to husband their programs and data; they regard them as products they have produced with investments of their own time and they have the rights to them (a given analysis can take 4-6 months; special data collection can take longer)

- 
- Relatively little replication takes place in economics for this reason
 - But also little replication takes place because there is little reward for a successful replication; journals will not publish it, at least not without more being done than simple replication
 - And, as in many other areas of scholarship, doing something new is more prestigious than examining someone else's work

- 
- NSF-economics has made some attempts to require data produced by economics NSF grants to be made available
 - NIH has been moving in this direction, and the social science side of NIH has been particularly proactive in pushing investigators to save and document their data publicly
 - Many other funders don't care, and most economic research is not supported by such funding in any case



Role of the Journals

- The situation makes the economics journals of particular importance in encouraging data sharing (and sharing of programs used to produce data and to conduct statistical analyses)
- The American Economic Review (official journal of the American Economic Association) has taken the lead on these issues


The AER


- About 1,300 submissions per year
- Most widely-read journal and most competitive
- Double-blind refereeing
- Submission fee
- Relatively wealthy journal because it has the financial support of the American Economic Association (20K members), unlike many other economics journals



The AER Data-Sharing Policy (Begun in 2005)

- (1) Researchers who submit articles for publication must clearly explain their data sources and procedures
- (2) Researchers whose articles are accepted for publication are required to provide both data and programs sufficient for replication, both of which are posted on the AER Web site

- 
- (3) Authors are encouraged to provide detailed data appendices which are posted on the AER Web site
 - (4) Researchers using proprietary data are required to supply their programs, and to provide instructions to other researchers on how to obtain permission to obtain the data
 - (5) The AER encourages the submission of replication/robustness papers

- 
- What the AER does not do:
 - (1) Does not require authors to supply data at time of submission, for reviewers; reviewers do not have the time to reanalyze the data and (thus far) authors are afraid that reviewers will steal their data and programs
 - (2) A fortiori, authors whose papers are not accepted for publication never have to supply anything




Results of the Policy


- Has been extremely well-received by the profession
- Most of the other leading journals in economics have adopted some similar type of policy
- Compliance has been 100 percent (maybe), although this is as a percent of non-proprietary data articles (about 10 percent of articles)



Future Issues

- Have researchers actually been in compliance in the sense of having posted the right data and programs in a usable form?
- Have not checked
- Are currently initiating a process of beginning to do so

- 
- Very few replication papers have been submitted
 - Should the journal take a more pro-active stance in encouraging (e.g., commissioning) replication/robustness studies, which would also provide a test of compliance?
 - Encouraging graduate students in economics departments to do this is another option

- 
- Some authors have complained about giving up data sets they have spent years in constructing
 - Economists call this the ‘patent’ problem
 - Have ignored the problem because there seems to be no practical way to address it



Conclusions

- Progress is being made, but it is slow
- Have a long way to go to get to the desired 'culture of replication' and data-sharing



Data Sharing in Economics and the Role of the American Economic Review

Robert A. Moffitt

Johns Hopkins University

Editor, American Economic
Review