Reproducibility Standards in Quantitative Social Science: Why Not Sociology?

Jeremy Freese
University of Wisconsin-Madison
and Harvard University

orienting considerations

- computers continue to increase the scale, sophistication, and extensiveness of quantitative analyses
- space constraints often require researchers to "paraphrase" or omit details of analyses
- increased concerns about the dependence of results on details of (*inter alia*) variable construction, case selection, model specification

Freese - Reproducibility Standards

orienting considerations

- Internet has revolutionized capacity for collective work
- materials previously available to anyone "upon request" can be made available to everyone at the time of publication
- many journals are adding possibilities for including online supplemental materials
- but why not use Internet to maximize transparency of procedures on which results are based?

Freese - Reproducibility Standards

3

economics data availability policy

AMERICAN ECONOMIC REVIEW

AEAweb: AER: Data Availability Policy

Aweb Journals Home ER Login Contacts Contents

Data Availability Policy

It is the policy of the American Economic Review to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the AER Web site. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the requirements above cannot be met.

Freese - Reproducibility Standards

Estimating the Value of Proposal Power

Brian Knight

Abstract | Full text for AFA members | Other only

Abstract | Full-text for AEA members | Other online viewing options

Link to Data Set [.zip]

Human Capital Formation, Life Expectancy, and the Process of Development

Matteo Cervellati and Uwe Sunde

<u>Abstract</u> | <u>Full-text for AEA members</u> | <u>Other online viewing options</u>

Link to Data Set [.zip]

The Impact of Outsourcing to China on Hong Kong's Labor Market

Chang-Tai Hsieh and Keong T. Woo

Abstract | Full-text for AEA members | Other online viewing options

Link to Data Set [.zip]

Using Experimental Economics to Measure Social Capital and Predict Financial Decisions

Dean S. Karlan

Abstract | Full-text for AEA members | Other online viewing options

Link to Data Set [.zip]

Manufacturer Liability for Harms Caused by Consumers

key points of economics data availability policy

- 1. data available at time of publication
- 2. archiving independent of author
- 3. "data" encompasses code and/or data
- 4. allowance made for "proprietary" data
- 5. makes provision of maximal information about analyses a mundane part of publication process

Freese - Reproducibility Standards

current sociology policy

- adapted from standard for psychology
- conceptualizes issue as individualistic and ethical matter
- ASA Code of Ethics 13.04(e): "once findings are publicly disseminated, sociologists permit their open assessment and verification by other responsible researchers"

Freese - Reproducibility Standards

7

advantages of social reproducibility policy

for suspiciously minded: reader asked to presume that sociologist would provide if asked the information economist has already provided up front

for less suspicious:

- presumption that good analysts are good archivists
- policy (and verifiability of findings) expires when researcher leaves discipline

Freese - Reproducibility Standards

potential objection #1: won't this imply more work for researchers?

- "implicit technical appendix" already assumed to exist
- policy would thus seem to add work only where more work would seem warranted anyway
- suggests anticipatory policy has benefits for individual practice beyond allowing opportunity for examination by others

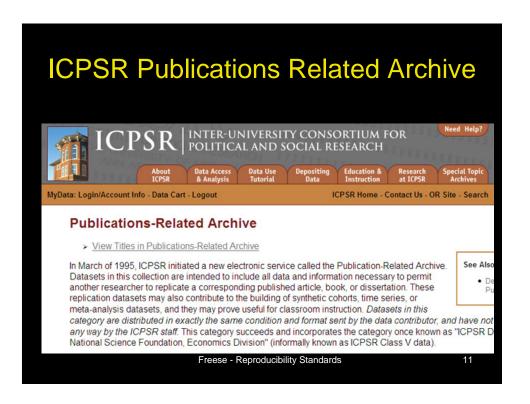
Freese - Reproducibility Standards

9

potential objection #2: won't this imply more work for editors?

- benefits of policy can be gained without requiring editors or their assistants to handle data
- "first footnote" discussing availability and depositing information in independent public archive

Freese - Reproducibility Standards



potential objection #3: there are good reasons for not making data available

Yes! For example:

- researcher does not have custodial right to the data
- researcher has custodial rights but is precluded from making data available for confidentiality reasons
- 3. researcher has custodial rights and does not regard sharing data as in intellectual interest

Freese - Reproducibility Standards

potential objection #3: there are good reasons for not making data available

- no argument against making data available is a good argument against providing explicit information about data availability
- wrong to believe that providing code for analysis is not worthwhile if data are not presently available

Freese - Reproducibility Standards

13

potential objection #4: there are good reasons for not making code available

- researchers may believe same interestprotection arguments for embargoing data should apply for code
- if so, raises question of how they think about their obligations about what they would provide "upon request"
- in any case, reviewers seem like they should have a right to assess the contribution implied by "proprietary" versus available code

Freese - Reproducibility Standards

key maxims, again

- 1. if researchers would make it available upon request, why not provide it now?
- 2. to whatever extent researchers cannot/will not make it available upon publication, why can't they be explicit about this?

Freese - Reproducibility Standards

15

conclusions

- many sociologists appear to have strange fatalism regarding possibility of change to their discipline's collective practices
- benefits for individual practice even if not collective practice
 - also benefits to signaling use of code or data sharing practice
- sociologists might consider how they may help to improve disciplinary practice through their work as reviewers

Freese - Reproducibility Standards