

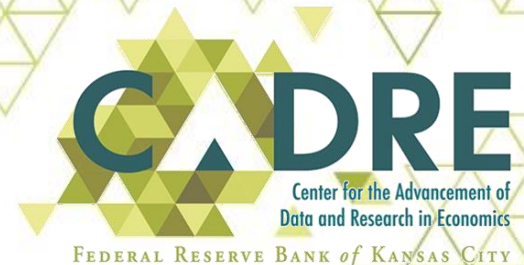


# You Can't Replicate What You Can't Find

## Data Preservation Policies in Economic Journals

Courtney R. Butler, MLS  
Brett D. Currier, MSLS, JD

\*The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.





# Background

# Copyright: Taylor & Francis

**THIS FORM IS A LEGALLY BINDING DOCUMENT. WE RECOMMEND THAT YOU RETAIN A COPY OF IT AND CONSULT A LEGAL ADVISOR IF YOU HAVE ANY QUESTIONS.**

## ASSIGNMENT OF COPYRIGHT: TERMS & CONDITIONS

### DEFINITION

1. Your article is defined as comprising (a) your Accepted Manuscript (AM) in its final form; (b) the final, definitive, and citable Version of Record (VoR) including the abstract, text, bibliography, and all accompanying tables, illustrations, data, and media; and (c) any supplemental material hosted by Taylor & Francis. This assignment and these Terms & Conditions constitute the entire agreement and the sole understanding between you and us ('agreement'); no amendment, addendum, or other communication will be taken into account when interpreting your and our rights and obligations under this agreement, unless amended by a written document signed by both of us.

### TAYLOR & FRANCIS' RESPONSIBILITIES

2. If deemed acceptable by the Editors of the Journal, we shall prepare and publish your article in the Journal. We may post your accepted manuscript in advance of the formal publication of the VoR. We reserve the right to make such editorial changes as may be necessary to make the article suitable for publication, or as we reasonably consider necessary to avoid infringing third-party rights or breaching any laws; and we reserve the right not to proceed with publication for whatever reason.
3. Taylor & Francis will deposit your Accepted Manuscript (AM) to any designated institutional repository including [PubMedCentral \(PMC\)](#) with which Taylor & Francis has an article deposit agreement, see 4.3. (c) below.

- xi. The right to expand your article into book-length form for publication provided that acknowledgement to prior publication in the Journal is made explicit (see below). Where permission is sought to re-use an article in a book chapter or edited collection on a commercial basis a fee will be due, payable by the publisher of the new work. Where you as the author of the article have had the lead role in the new work (i.e., you are the author of the new work or the editor of the edited collection), fees will be waived. Acknowledgement to prior publication in the Journal should be made explicit (see below):

**Acknowledgement:** This <chapter or book> is derived in part from an article published in <JOURNAL TITLE> <date of publication> <copyright Taylor & Francis>, available online: <http://www.tandfonline.com/> <Article DOI>

If you wish to use your article in a way that is not permitted by this agreement, please contact [permissionrequest@tandf.co.uk](mailto:permissionrequest@tandf.co.uk)

### WARRANTIES MADE BY YOU AS AUTHOR

5. You warrant that:
  - i. All persons who have a reasonable claim to authorship are named in the article as co-authors including yourself, and you have not fabricated or misappropriated anyone's identity including your own.

# Responsive to the Community

- Comprehensive list of journals
- 10+ year half life
- Local data preservation platform

# What does reproducible mean?

Type	Definition
Replication	Estimating parameters drawn from the same sampling distribution as those in the original study
Verification	Ensuring that the exact statistical analysis reported in the original paper gives materially the same results reported in the paper, either using the original dataset or remeasuring with identical methods the same traits of the same sample of subjects.
Reproduction	Resampling precisely the same population but otherwise using identical methods to the original study.

Clemens, Michael A. "The meaning of failed replications: A review and proposal." *Journal of Economic Surveys* (2015).



# What does reproducible mean?

Type	Definition
Robustness	Estimating parameters drawn from a different sampling distribution from those in the original study
Reanalysis	Altering the computer code from the original study.
Extension	Using new data—gathered on a sample representative of a different population, or gathered on the same sample at a substantially different time, or both.

Clemens, Michael A. "The meaning of failed replications: A review and proposal." *Journal of Economic Surveys* (2015).

# What does reproducible mean?

Type	Definition
Exact Reproduction	Reproducing the data exactly as presented in the paper

# Theoretical reproducibility & item type

Type	Example Files	Tentative Use
Verification	Original Dataset Original Computer Code Original Methods	Original Dataset (or near identical) Original statistical methods
Reproduction	Original Computer Code Original Methods	Original Population Identical methods
Reanalysis	Original Computer Code Original Dataset	Altered Computer Code Identical Data
Extension	Original Computer Code Original Dataset	Identical Computer Code New data
Exact Reproduction	Chart file	Reproduced in the review section



# Graphics

Blake, Judith. "Are babies consumer durables? A Critique of the Economic Theory of Reproductive Motivation. Population Studies 22, no. 1 (1968): 5-25.

## ARE BABIES CONSUMER DURABLES?

9

Even so, if one looks up the one article he cites, one finds that even the statistical regularities are open to question.<sup>17</sup>

*Subscribers to a Consumers' Union.* Becker also utilizes some previously unpublished data on 'family income, education, earners, and dependent children of a sample of the subscribers to Consumers' Union'.<sup>18</sup> This group, he claims, 'is particularly valuable for our purposes since it primarily consists of families with a keen interest in rational, informed consumption. If my analysis is at all relevant, fertility and income should be more positively related in this group than in the U.S. population as a whole.'<sup>19</sup>

The figures he gives are reproduced in Table 1. From them Becker concludes that the 'income elasticity is about 0.09 and 1.14 for graduates of a four-year college and of a graduate school respectively. These data, then, are very consistent with my analysis, and indicate that well-informed families do have more children when their income increases.'<sup>20</sup>

TABLE 1. *Reproduction of Becker's table on average number of dependent children for single-earner families with head age 35-44 in a sample of subscribers to Consumers' Union, April 1958\**

Income class	Average number of dependent children by education class of head			
	High school graduate – or less	Some college	Graduate of four-year college	Graduate degree
Less than \$3,000	2.43	1.61	2.50	2.17
\$3,000– 3,999	2.14	2.47	2.18	2.23
4,000– 4,999	2.70	2.40	2.04	2.18
5,000– 7,499	2.68	2.73	2.88	2.67
7,500– 9,999	2.80	2.94	3.00	3.03
10,000–14,999	2.89	3.03	3.12	3.23
15,000–24,999	2.85	3.04	3.04	3.31
25,000 and over	3.12	3.23	3.28	3.60

\* Reproduced from Becker, *loc. cit.*, p. 221.

<sup>17</sup> Lamson's data on family size were obtained from students who reported on births and child mortality in their families. He himself admits that: 'It is quite likely that in these families a certain number of miscarriages, stillbirths, and abortions have been omitted through failure of the reporting students either to know or to state such facts concerning the puerperal history of their mothers' (pp. 308–309). Lamson assumes, however, that these errors are randomized over the different groups he is studying. But, this would only be true if the lower economic groups were no more subject to such risks than the upper economic groups. Data such as these cannot take into account the role of female infanticide, or the sale of infant girls into prostitution. Moreover, many of the findings are themselves highly suspect. For example, the highest fertility of all is found among the higher-educated Christian families, yet Lamson does not tell us whether these families were Catholic or not. If many of them were, their behaviour (as converts) would include neither contraception nor infanticide. In addition, although Lamson examines concubinage in relation to fertility (and finds that concubinal families do not have the highest reproduction), he does not bring out that concubinage was concentrated at the upper-income levels. He thus overlooks the fact that upper-income, non-Christian families had at their disposal a hedge against wifely infertility or sub-fecundity. The average fertility of such families could be kept high through the reproductive services of concubines.

<sup>18</sup> Becker, *op. cit.*, p. 221.

<sup>19</sup> *Ibid.*, p. 221.

<sup>20</sup> *Ibid.*, p. 222.



# Project and Method





# Project

## **Question:**

What is the industry standard regarding data availability and data preservation in economic journals?

## **Method:**

Survey top journals from economics and related disciplines to determine current practice



# Data Availability or Replication Policy

Present	Not Present
<ul style="list-style-type: none"><li>• Journal mandated or strongly suggested underlying data be made available, either as a direct submission or upon request</li></ul>	<ul style="list-style-type: none"><li>• No mention of data submission</li><li>• Optional to include supplementary data (not required or recommended)</li></ul>

# Data Preservation Policy\*

Present	Not Present
<ul style="list-style-type: none"><li>• Included a discussion of long-term preservation outside of a data availability or replication policy with at least one of the following:<ul style="list-style-type: none"><li>• Assignment of a DOI or other persistent identifier</li><li>• Restrictions on file formats, size, or specifications</li><li>• Requirement of preservation or contextual information/metadata</li><li>• Deposit in an internal repository with preservation features</li></ul></li></ul>	<ul style="list-style-type: none"><li>• Grouped in with an overarching data availability/replication policy</li><li>• Burden of preservation placed on the author without guidance</li><li>• Submission 'instructions' not considered a preservation policy unless preservation was clearly stated as the purpose for the requirements</li></ul>

\*Includes tables, figures, and other paper components containing data




# Population and Sampling

- **Academic, peer-reviewed journals that publish content on economics and related disciplines**
- **Purposeful selection based on three common ranking systems and one measure of local activity:**
  - Web of Science InCites Journal Citation Reports
  - Google Scholar Metrics
  - IDEAS/RePEc Aggregate Rankings for Journals (All Years)
  - Federal Reserve Bank of Kansas City Research Staff CVs



# Subset

- Top 250 journals from combined select Web of Science categories  
(250 total)
  - Top 20 journals from individual select Google Scholar Metrics categories  
(271 total after de-duplication)
  - Top 250 journals from IDEAS/RePEc  
(250 total)
  - All journals with at least one FRBKC interaction (publication or service)  
(152 total)
  - **923 journals in subset**
- 



# Select Categories

## Google Scholar Metrics


- Accounting & Taxation
- Business, Economics & Management (general)
- Development Economics
- Economic History
- Economic Policy
- Economics
- Entrepreneurship & Innovation
- Finance
- Game Theory & Decision Science
- Human Resources & Organizations
- International Business
- Marketing
- Strategic Management
- Tourism & Hospitality
- Probability & Statistics with Applications

## Web of Science

- Agricultural Economics & Policy
  - Business
  - Business, Finance
  - Economics
  - Industrial Relations & Labor
  - Management
  - Statistics & Probability
- 




# Additional Selection

- Cross-referenced all four sources, included any journal that appeared more than once  
(220 journals)
  - Added remaining journals in top 50 from Google Scholar Metrics, IDEAS/RePEc, and Web of Science  
(20 journals)
  - Added remaining journals with 2+ FRBKC interactions  
(5 journals)
  - **245 journals before exclusions**
- 




# Exclusions

- 9 exclusions
    - Journals that were not peer-reviewed
    - Journals that did not accept articles relating to data/economics and related disciplines
    - Journals no longer in print
    - Federal Reserve publications
  - **236 journals in final sample**
- 





# Text Analysis

- Identified and downloaded journal policies, author instructions or guidelines, and style guides from journal website(s)
  - Analyzed text to identify inclusion in 2 categories:
    - Data or Replication Policy
    - Data Preservation Policy
- 



# Results

- 80 journals (34%) have a data availability or replication policy
- 0 journals (0%) have a data preservation policy

*Re-analyzing text to identify guidelines consistent with preservation features*



# Guidelines Consistent with Preservation Standards\*

Included suggestions or requirements for at least one of the following:

- Assignment of a DOI or other persistent identifier
- Restrictions on file format, size, or specifications
- Preservation or contextual information/metadata
- Deposit in any repository with preservation features

Included even if these features were entirely optional or not part of a clearly defined “policy”

\*Includes tables, figures, and other paper components containing data



# Preliminary Results

181 journals (77%) have guidelines consistent with preservation standards

Of the 80 journals with a data availability or replication policy, 76 journals (95%) also have guidelines consistent with preservation standards





# Conclusions



# Conclusions

- Data availability policies are becoming more common but still not standard
- Most journals have at least one feature in guidelines that are consistent with preservation standards
- Reproducibility cannot truly be achieved until a greater focus is placed on long-term preservation

# Implications & Future Directions

- Author Publication Guidance needs to extend beyond the article
- Working with journals
  - To retain the copyright OR
  - Link to data archive
- Data Archives



Questions?