Data policies and data archives: A new paradigm for academic publishing in economic sciences?

Sven Vlaeminck | Leibniz Information Centre for Economics (ZBW)
2015/9/2 | 19th International Conference on Electronic Publishing (ELPUB 2015) | Valletta | Malta

Funded by

EDaWaX
European Data Watch Extended

Source: Economics-E-Journal
Source: American Economic Review
Source: Dataverse-Network / ckan.org

The Welfare Effects of Bundling in Multichannel Television Markets (pp. 643-85)
Gregory S. Crawford and Ali Yurukoglu
Abstract/Tools | Full-Text Article | Download Data Set | Online Appendix

The Use of Full-Line Forcing Contracts in the Video Rental Industry (pp. 666-719)
Katherine Ho, Justin Ho and Julie Holland Mortimer
Abstract/Tools | Full-Text Article | Download Data Set | Online Appendix

Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World (pp. 720-49)
Dean Fudenberg, David G. Rand and Anna Dreber
Abstract/Tools | Full-Text Article | Download Data Set | Online Appendix

The Dataverse Network Project

Source: Dataverse-Network / ckan.org
Table of content:

> 1. Introduction & motivation
> 2. The study
  – 2.1 Approach and methodology
  – 2.2 Characteristics of our research sample
> 3. Some results of our study
  – 3.1 The number of journals equipped with data policies
  – 3.2 Some specifications these data policies
> 4. Conclusion and 'take home messages'
1. Introduction & Motivation
Changing patterns in economic research

> Hamermesh analysed 748 articles in economic top journals (published between 1963 and 2011).
> He found significant changes in the methodology used by researchers:
  - Studies based on borrowed and own data sets, experimental approaches and simulations comprise roughly 80% of all surveyed articles in 2011.*
  - In 1963 this percentage was below 50%.*

Challenges in empirically-based research


> 2012/2013: EU-Monetary Affairs Commissioner Rehn and US vice presidential candidate Ryan used the results of Reinhart & Rogoff to justify austerity policy.

> 2013: A student found multiple failures in the calculations and selectively omitted data.

⚠️ Strong doubts about Rogoff’s & Reinhart’s findings
How do journals in economic sciences react to these challenges?

> Do we find a large number of journals equipped with suitable data policies?

> And how are these policies structured? What type of replication files and supplemental information do they request?

> Do these data policies really foster replicable research?

> Can we even speak of a new paradigm for publishing in economic sciences?
2. The Study

> 2.1 Approach & methodology
> 2.2 Some characteristics of our research sample
Approach & methodology:

> Creation of a research sample based on two lists of journals provided by German learned societies (in economics & business studies).
> We carefully checked the webpage(s) of each journal to determine whether it has a data policy or not.
> We analysed the specifications of each policy we found.
> We compared the specifications of each policy to some requirements for reproducible research*

What we checked

> What do authors have to provide to satisfy the policy?
  – dataset(s) employed?
  – code of computation (syntax)?
  – self-compiled software components?
  – descriptions of the data (data dictionary/codebook)?

> Is the policy mandatory?

> Who is responsible to provide the data (author/journal?)

> When do authors have to submit their replication data?

> Does the policy have a procedure to foster replication for proprietary/confidential datasets?

> ...
2. The Study

> 2.1 Approach & methodology
> 2.2 Some characteristics of our research sample
Primary subject area (n=346)

Primary Subject Area of Journals in our Sample (n=346)

- 160; 46.2% primarily business studies
- 132; 38.2% economics & business sciences in equal parts
- 34; 9.8% primarily economics
- 20; 5.8% other
Distribution of our sample by rating in the Handelsblatt ranking 2012

(n=325)

<table>
<thead>
<tr>
<th>n</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.7</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18</td>
<td>55</td>
<td>40</td>
<td>47</td>
<td>77</td>
<td>66</td>
<td>22</td>
</tr>
</tbody>
</table>

Distribution of the full sample
3. Some results of our study

> 3.1 The number of journals equipped with data policies
> 3.2 Specifications of journal’s data policies
Data Policies im Sample:

Data Policies in full Sample
(n=346)

- Data Availability Policy: 14.2% (49)
- Author Responsibility 'Policy': 6.4% (22)
- No Policy: 79.5% (275)
Economics & Business studies: Strong differences

Data Policies of Journals in Economics and Business Studies
(n=293 | Econ: n=132; Bus.Sc.: n=160)

- Economics:
  - Data availability policy: 25.8%
  - Author responsibility 'policy': 6.1%
  - No policy: 68.2%

- Business Studies:
  - Data availability policy: 6.9%
  - Author responsibility 'policy': 5.6%
  - No policy: 87.5%
Data policies and journals’ rankings

Ranking of journals by type of data policy [median]

- Handelsblatt Ranking Business Studies 2012 (n=325)
- Thomson Reuters Impact Factor® (n=262)
- Jourqual 2.1 Ranking (2011) (n=257)

Legend:
- Green diamond: Data Av. Policies
- Yellow square: Author Resp. Policy
- Red triangle: No Policy
3. Some results of our study

> 3.1 The amount of journals equipped with data policies
> 3.2 Some specifications of journal’s data policies
Requirements of journals’ data policies

Data Policies: What authors have to provide

\( n=71; 49 \text{ Data Av. Policies } & 22 \text{ Author Responsibility Policies} \)

- **Datasets**: 100.0%
- **Code/Script**: 77.6%
- **Software/Programs**: 53.1%
- **Descriptions**: 71.4%
- **Intermediate Datasets**: 28.6%

EDaWaX
European Data Watch Extended
The share of mandatory policies

Mandatory Data Policies in Economic Science Journals

(n=71)

Data Availability Policies

61.2%

Author Responsibility Policies

40.9%
The case of proprietary data

Data Policies: Defined Procedures for Research based on proprietary Data
(n=68)

- Data Availability Policy: 52.2%
- Author Responsibility Policy: 9.1%
Results: Data policies‘ specifications

> ARPs define far less policy specifications than DAPs = weak policies / „window dressing“
> The code of computation/syntax is requested by (too) few data policies. The same applies to self-compiled software.
> 60% of all DAPs are mandatory; not really a satisfactory result.
> Unfortunately, only half of all DAPs have a procedure to forster replicability also for proprietary/confidential data.
> Intermediate datasets are required by NONE of the policies; less than a third of all DAPs invites authors to submit these files.
4. Take home messages

> To date, journals equipped with useful data availability policies still are a minority.

> But, especially in economics the share of journals equipped with such policies rises (> 25% in our sample).

> On average, journals with such policies are among the best rated journals in the field.

> Among data policies, there is light and shadow. To date, many data policies remain weak. But: A „good practise“ policy has been implemented by the American Economic Review (AER).

It is too early to say, that data policies and archives have become a new paradigm for publishing in economic sciences.
Thank you very much for your attention!

...do you have questions or comments?

Contact:
Sven Vlaeminck | s.vlaeminck@zbw.eu | www.edawax.de
ZBW – Leibniz Information Centre for Economics
Neuer Jungfernstieg 21
D- 20354 Hamburg | Germany