

# Data and Replication Policies in Economics

Arthur Seibold (University of Mannheim)

Open Science Day, Mannheim, 12 October 2023

# Introduction

- Long-standing trend towards empirical research in economics (cf. Hamermesh, 2013)
  - 1980s: 58% of papers published in top journals purely theoretical, 42% with some empirical analysis
  - 2010s: 19% pure theoretical, 81% with empirics
- Critique: lack of transparency and replicability in empirical research
  - Description of methods in published paper typically not sufficient
- In recent years, many economics journals have implemented **Data and Replication Policies**

# Data and Replication Policies in Economics Journals

- Pioneered by the American Economic Association (AEA)
  - Initial policy (2005-2019) required posting analysis codes + data
  - Current policy requires posting full codes + data
  - Since 2018: data editor oversees compliance and leads reproducibility check
  - Journals covered: *American Economic Review* + all other AEA journals
- By now, all other top economics journals have data/replication policies
  - *Review of Economic Studies* has policy following DCAS + data editor (since 2019)
  - *Econometrica* has own policy + data editor (since 2023)
  - *Quarterly Journal of Economics* and *Journal of Political Economy* follow AEA policy (but no data editor)
- Increasing number of journals are following, but policies still vary

# The AEA Data and Code Availability Policy

- Papers can only be published if authors provide
  1. Data sets
  2. Description sufficient to access all data at original source location
  3. Programs used to create final data from raw data
  4. Programs to run analysis
  5. Description sufficient to run programs
    - full replication possible
- If data proprietary, need to provide information on how and where data can be accessed
- Data editor (or third-party replicator) conducts reproducibility check
- Replication package archived in AEA data and code repository

## Practical Experience: Seibold (AER 2021)

- Paper studies why so many people retire at Normal Retirement Age in Germany (it is not financial incentives!)
- Timeline
  - I submitted to *American Economic Review* in July 2019
  - First editorial decision (revise and resubmit) in November 2019
  - I resubmitted in April 2020
  - Editor decided to *conditionally accept* in June 2020
  - I submitted replication package in August 2020
  - Data editor requested revisions to replication package in October 2020
  - Paper was finally accepted in December 2020

## Practical Experience: Seibold (AER 2021)

- Putting together replication package took a while
  - Organize codes
  - Run full replication (required one-week visit at German State Pension Fund)
  - I detected one small error in results → notify editor
  - Carefully check data access modalities
  - Write README document
  - Include appropriate data citations in paper
- Due to confidential data access, only manual reproducibility check was performed
- However, my initial replication package lacked some required elements → revisions required

## Report from Data Editor

[Image: title page of data report]

# Report from Data Editor

Revisions requested

[Image: list of requested revisions]



# Report from Data Editor

Contacted data providers

[Image: report on contact with data providers]

# Report from Data Editor

Some data sets missing

[Image: requested revisions, specifically on providing data sets]

# Report from Data Editor

“Manual” reproducibility check

[Image: reproducibility check, list of codes with results and data sets checked]

# Report from Data Editor

“Manual” reproducibility check

[Image: reproducibility check, list of codes with results and data sets checked]

## Conclusion

- Leading economic journals have implemented data and replication policies
- Large benefits
  - Make data sets and coding methods widely available
  - Full transparency on empirical methods
  - Reproducibility check directly performed
  - Permit replication by other researchers
- But also costs
  - Journals have to employ data editors with large teams
  - Authors have to spend considerable time and effort
  - Particularly costly to junior researchers