The Reproducibility of Economics Research: A Case Study

S.Herbert H.Kingi F.Stanchi and L.Vilhuber

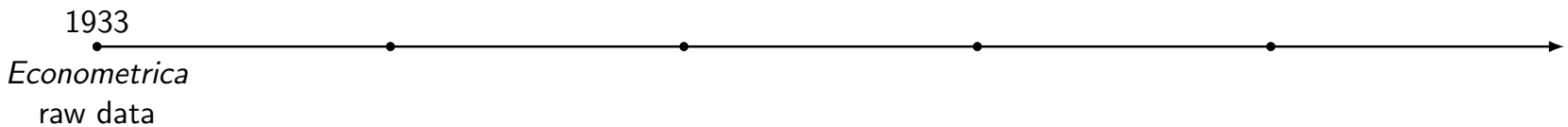
November 10, 2021

Credibility of research in economics

- Transparency and openness are essential to the credibility of economics research
- Even more important for policy making because research provides informs economic policies
- Evidence-based policy making with reliance on academic research
- Research needs to be trustworthy, hence reliable

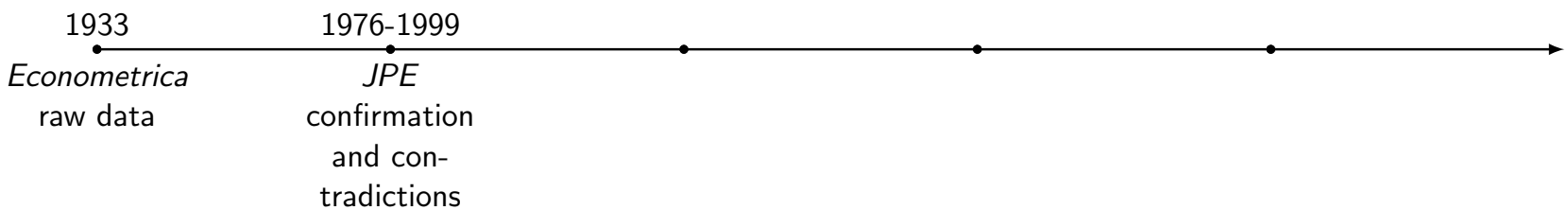
Peer review and reproducibility

- Peer review ensures high quality and original research
- But zero obligation to reproduce results from the codes and data
- Data availability policies to promote transparency



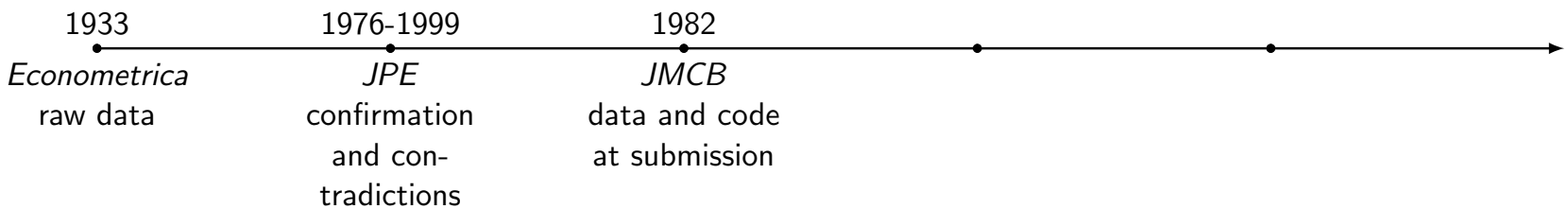
Peer review and reproducibility

- Peer review ensures high quality and original research
- But zero obligation to reproduce results from the codes and data
- Data availability policies to promote transparency



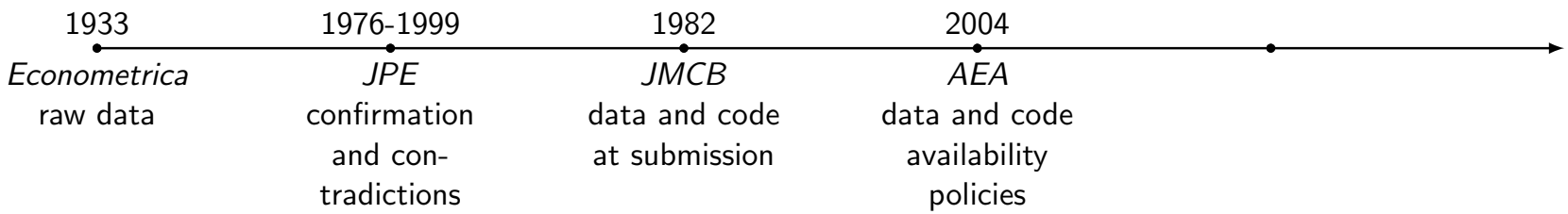
Peer review and reproducibility

- Peer review ensures high quality and original research
- But zero obligation to reproduce results from the codes and data
- Data availability policies to promote transparency



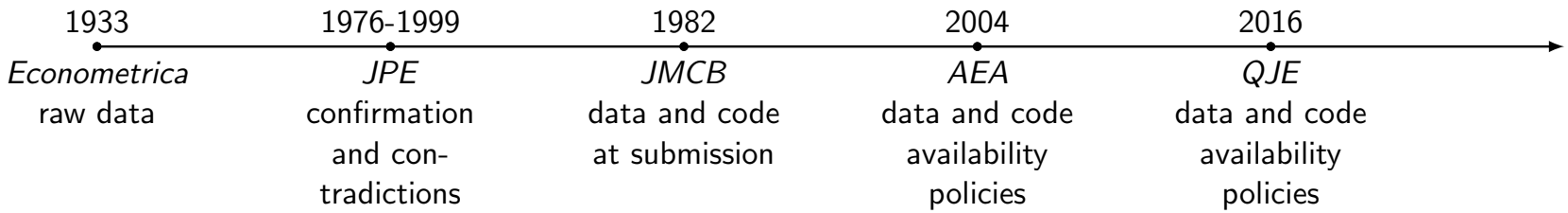
Peer review and reproducibility

- Peer review ensures high quality and original research
- But zero obligation to reproduce results from the codes and data
- Data availability policies to promote transparency



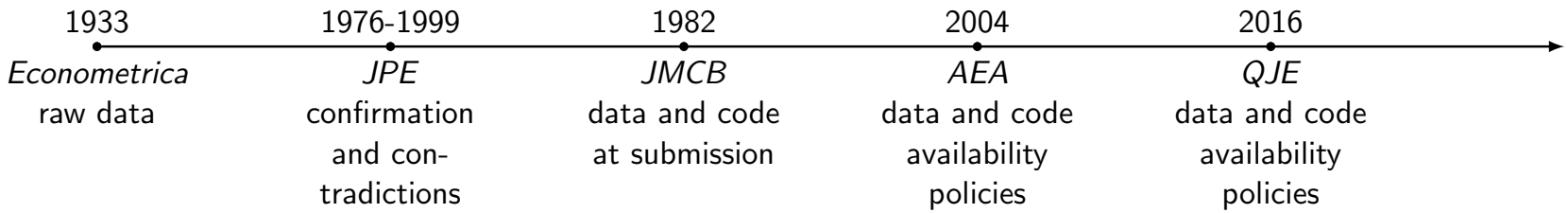
Peer review and reproducibility

- Peer review ensures high quality and original research
- But zero obligation to reproduce results from the codes and data
- Data availability policies to promote transparency



Peer review and reproducibility

- Peer review ensures high quality and original research
- But zero obligation to reproduce results from the codes and data
- Data availability policies to promote transparency



- Between 8.1 % and 54% of economics journal have it, depending on sample (Duvendack et al., 2015; Vlaeminck and Herrmann, 2015; Hoffler 2017)

What we do

- AEA data policy enforced in 2005

  "Authors of accepted papers ... must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. ... 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."

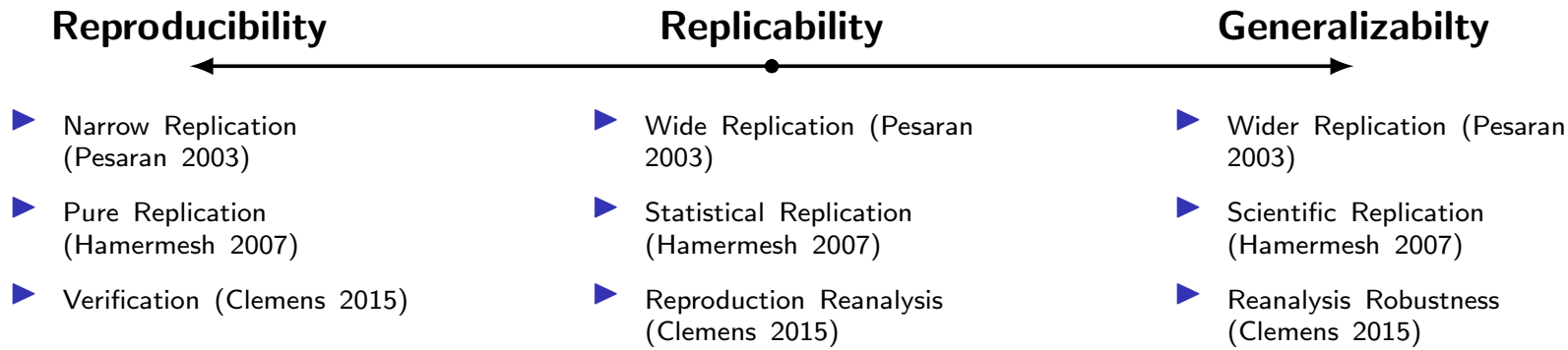
What we do

- AEA data policy enforced in 2005

  "Authors of accepted papers ... must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. ... 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."

- **Can data availability policies with light enforcement yield reproducible research?**

Variety of replication concepts



- Reproducible research: same materials + same procedure as original investigator = duplicate results

- Can undergraduates, **armed only with the information provided by authors on the journal website, successfully reproduce the tables and figures** presented by the author in the article?

Preview of results

Data availability policies (unverified reproducibility) yields reproducible research:

1. **Moderate replication success:** 28% (total articles), 37% (total assessed), 42% (non-confidential data)

2. **Main reason is confidential data, with the supplementary material complying with the policy overall**
   - Documentation complete (73% provided complete readme)
   - A proportion successful replications still required some code changes
3. **No citation bonus** for clean and ready to use research

- Results emphasize the need for a systematic check of reproducibility by journals (AEA policy post 2018), or third-party (Cascad)

Reproduction procedure

Reproduction procedure

Replication lab: undergraduates over 5 summers (2014-2018)

- Minimum knowledge: economics, computer sciences or Operations Research
- Training: version control systems, cloud computing, reproduction procedure

3-step process:

1. Entry questionnaire
    - Descriptive information: online appendix, data sets presence and availability, programs, documentation and its clarity
    - Expected level of replication difficulty

Reproduction procedure

Replication lab: undergraduates over 5 summers (2014-2018)

- Minimum knowledge: economics, computer sciences or Operations Research
- Training: version control systems, cloud computing, reproduction procedure

3-step process:

1. Entry questionnaire
    - Descriptive information: online appendix, data sets presence and availability, programs, documentation and its clarity
    - Expected level of replication difficulty
2. Reproduction exercise: record any change in VCS, replication summary in a txt

Reproduction procedure

Replication lab: undergraduates over 5 summers (2014-2018)

- Minimum knowledge: economics, computer sciences or Operations Research
- Training: version control systems, cloud computing, reproduction procedure

3-step process:

1. Entry questionnaire
    - Descriptive information: online appendix, data sets presence and availability, programs, documentation and its clarity
    - Expected level of replication difficulty
2. Reproduction exercise: record any change in VCS, replication summary in a txt
3. Exit questionnaire
    - Reproduction success, reasons for failure

Reproducing the AEA:applied journal

- 303 papers of the journal *AEJ:applied economics* examined
- Covering years well after introduction of the data availability policy: 2009-2018

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
23	32	36	40	10	40	24	36	42	20	303

Compliance with data availability policy

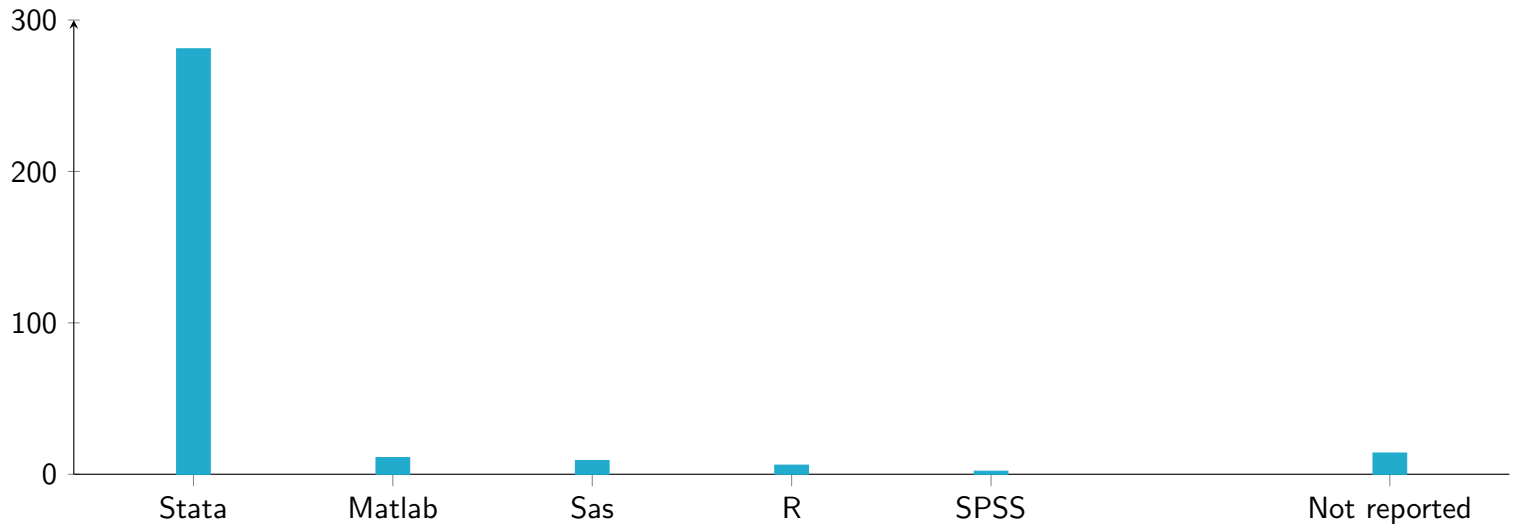
Compliance with AEA data provision requirements

- Not all articles were accompanied by the necessary data set

Reason	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total	Percent
Confidential Data	5	10	8	10	1	15	2	11	11	7	80	26.4
Data was Provided	16	22	28	27	9	23	21	23	29	11	209	68.98
No Data or Reason	2			3		2	1	2	2	2	14	4.62
Total	23	32	36	40	10	40	24	36	42	20	303	100

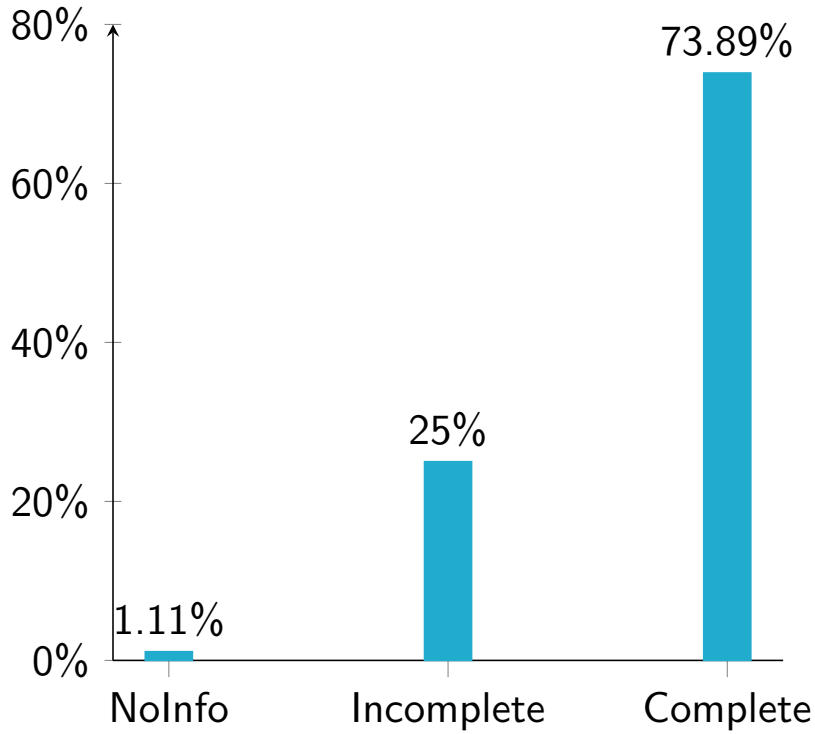
- Mainly due to confidential data
- Imperfect provision of data, with missing data: "No data or reason"

Dominance of stata



Compliance with AEA data availability policy

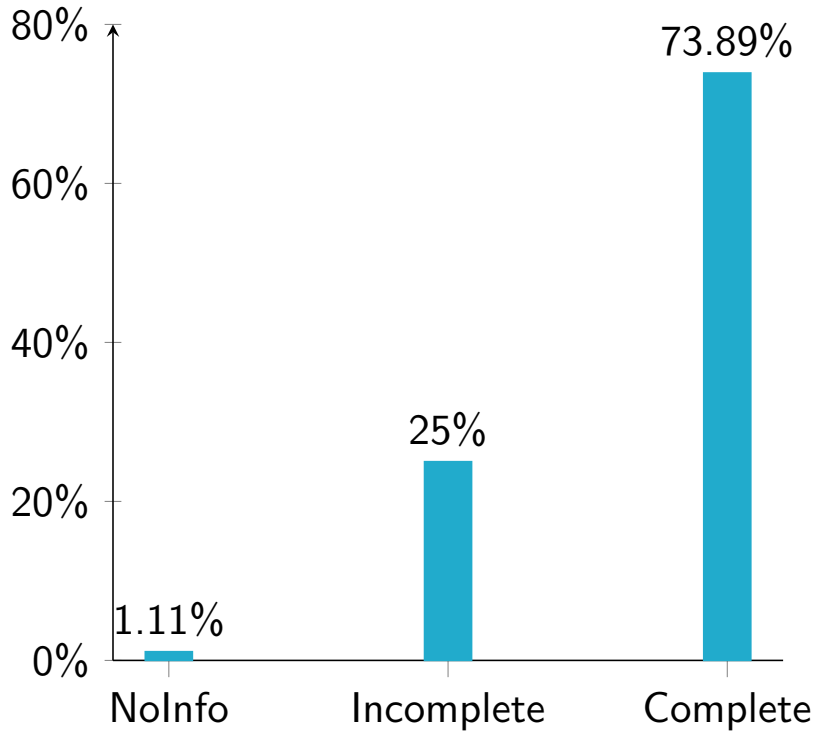
- Documentation quite complete: majority of articles is well-documented



- Complete: instructions about how to replicate

Compliance with AEA data availability policy

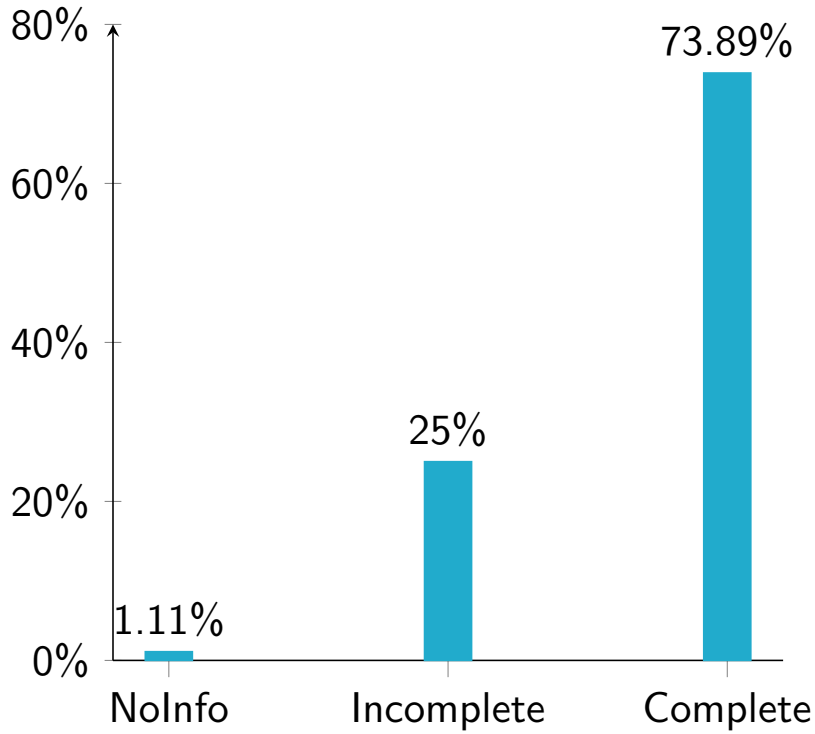
- Documentation quite complete: majority of articles is well-documented



- Complete: instructions about how to replicate
- Incomplete: ambiguous ReadMe file

Compliance with AEA data availability policy

- Documentation quite complete: majority of articles is well-documented



- Complete: instructions about how to replicate
- Incomplete: ambiguous ReadMe file
- No Info: missing ReadMe file

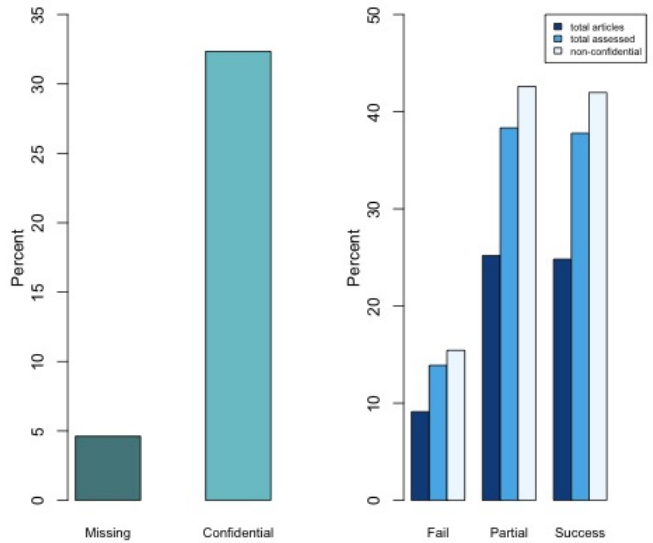
Subjective measure of replication difficulty

Difficulty Rating	Number of Articles	Percent
1	64	21.12
2	64	21.12
3	66	21.78
4	37	12.21
5	72	23.76

- Difficulty level 1
  - Data provided & public
  - Documentation clear & complete
  - Code ready to run
- Difficulty level 3
  - Documentation unclear
  - Code order ambiguous
- Difficulty level 5
  - Missing Data/Code/Documentation

Can data availability policies ensure reproducibility?

Moderate replication success



- **Full replication success** goes from
  - 25% (missing, confidential and data eligible for reproduction)
  - to 38% (180 assessed including confidential data identified during the reproduction)
  - to **42**% conditional on non-confidential data
- **Partial success**: another **42**%

Replication success

Comparable with other exercises using different journals and methodologies:

- 13% Dewald et al. (1986)
- 6% McCullough et al. (2006)
- 43% Chang and Li (2015)
- 61% Camerer et al. (2016)

Main reasons for unsuccessful replication

Year	Missing Data	Corrupted Data	Code Error	Software Unavailable	Other
2009		1			
2010					3
2011					12
2012	1	1	1		
2014	1			1	
2016	1				
2017	1				1
2018		1			
Total	4 (16%)	3 (16%)	1 (4%)	1 (4%)	16 (64%)

1. Confidential and proprietary data: 80 identified during the initial assessment and 18 during reproduction

Main reasons for unsuccessful replication

Year	Missing Data	Corrupted Data	Code Error	Software Unavailable	Other
2009		1			
2010					3
2011					12
2012	1	1	1		
2014	1			1	
2016	1				
2017	1				1
2018		1			
Total	4 (16%)	3 (16%)	1 (4%)	1 (4%)	16 (64%)

1. Confidential and proprietary data: 80 identified during the initial assessment and 18 during reproduction

2. Inconsistent values

Main reasons for unsuccessful replication

Year	Missing Data	Corrupted Data	Code Error	Software Unavailable	Other
2009		1			
2010					3
2011					12
2012	1	1	1		
2014	1			1	
2016	1				
2017	1				1
2018		1			
Total	4 (16%)	3 (16%)	1 (4%)	1 (4%)	16 (64%)

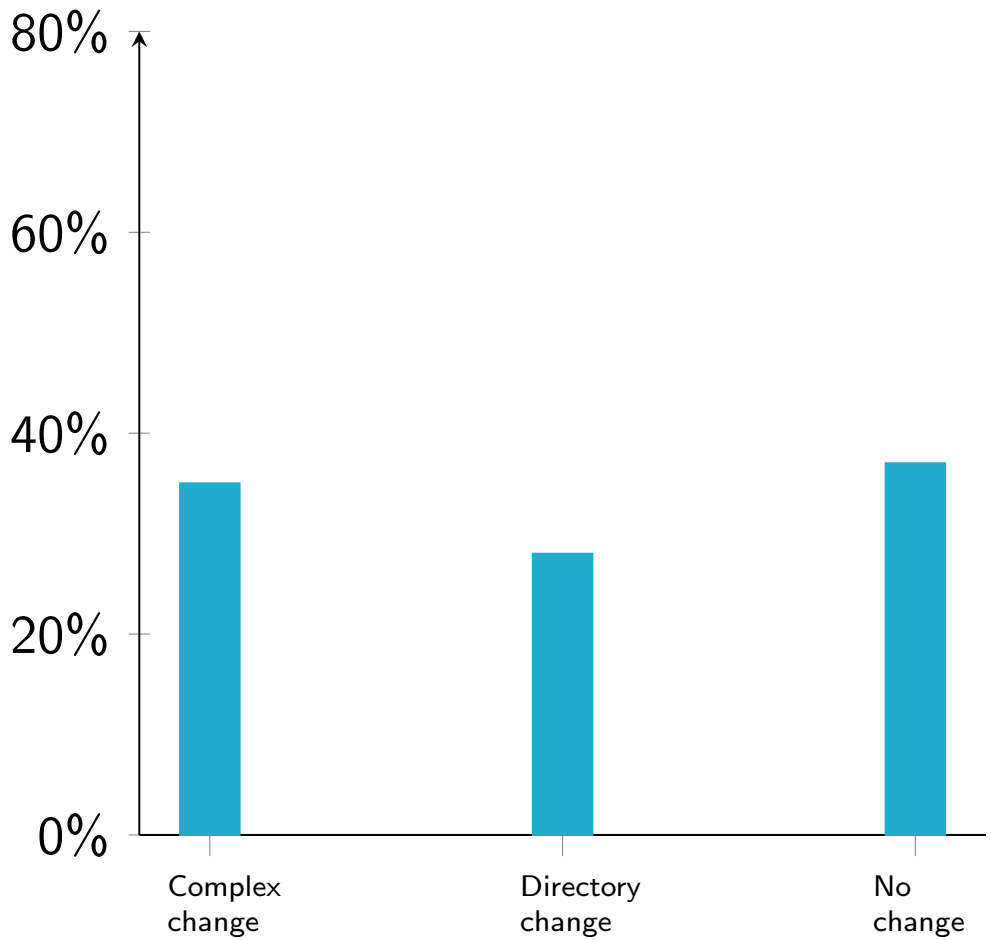
1. Confidential and proprietary data: 80 identified during the initial assessment and 18 during reproduction

2. Inconsistent values

3. Missing or corrupted data

Successful replications codes

- Majority required at least a directory change
- 1/3 of successful papers required complex code modifications



Good documentation is positively associated with reproductive success

	Successful Reproduction
Documentation Clarity = Complete	0.300***
	(0.080)
Constant	0.200**
	(0.070)
N	180
Notes:	Dependent variable = 1 if fully replicated.

- Confirming Findings/Suspicions of Chang and Li (2015), McCullough et al. (2006) and Stark (2018)

Is there a citation bonus for reproducible papers?

Incentives to provide reproducible material?

	Annual Citations					
	(1)	(2)	(3)	(4)	(5)	(6)
avghindex	0.200***			0.200***		
	(0.030)			(0.060)		
tophindex		0.070***			0.070**	
		(0.020)			(0.030)	
lowhindex			0.100**			0.090
			(0.050)			(0.080)
'Fully reproduced'	0.600*	0.500	0.500			
	(0.400)	(0.300)	(0.400)			
avghindex:'Fully reproduced'	−0.070					
	(0.050)					
tophindex:'Fully reproduced'		−0.040				
		(0.030)				
lowhindex:'Fully reproduced'			−0.070			
			(0.080)			
'Full or Partial'				0.400	0.100	−0.100
				(0.500)	(0.400)	(0.500)
avghindex:'Full or Partial'				−0.060		
				(0.070)		
tophindex:'Full or Partial'					−0.020	
					(0.030)	
lowhindex:'Full or Partial'						0.000
						(0.090)
Constant	2.000***	2.000***	2.000***	2.000***	2.000***	3.000***
	(0.200)	(0.200)	(0.300)	(0.500)	(0.400)	(0.400)
Observations	78	78	78	78	78	78

Positive but noisy citation bonus for confidential data

	Annual Citations		
	(1)	(2)	(3)
avghindex	3.000***		
	(0.700)		
tophindex		1.000***	
		(0.400)	
lowhindex			2.000
			(1.000)
confidential_data	17.000*	13.000*	8.000
	(9.000)	(8.000)	(9.000)
avghindex:confidential_data	−1.000		
	(1.000)		
tophindex:confidential_data		−0.500	
		(0.700)	
lowhindex:confidential_data			−0.200
			(2.000)
Constant	5.000	13.000***	20.000***
	(6.000)	(5.000)	(5.000)
N	119	119	119

Conclusion

- Moderate replication success even with data availability policies (DAP)
- Main culprit is confidential data, so good compliance overall with the DAP
  - online material quality is not yet perfect to reach reproducibility (some changes in the code were still required for successful reproductions)
  - non-zero proportion of average quality documentation
- Incentives to provide reproducible research not in the citation bonus

How do we reach the "good" equilibrium?

- Systematic verification (AEA data editor 2018) leads to higher re-execution (Trisovic, 2021)
- Third-party services such as cascad