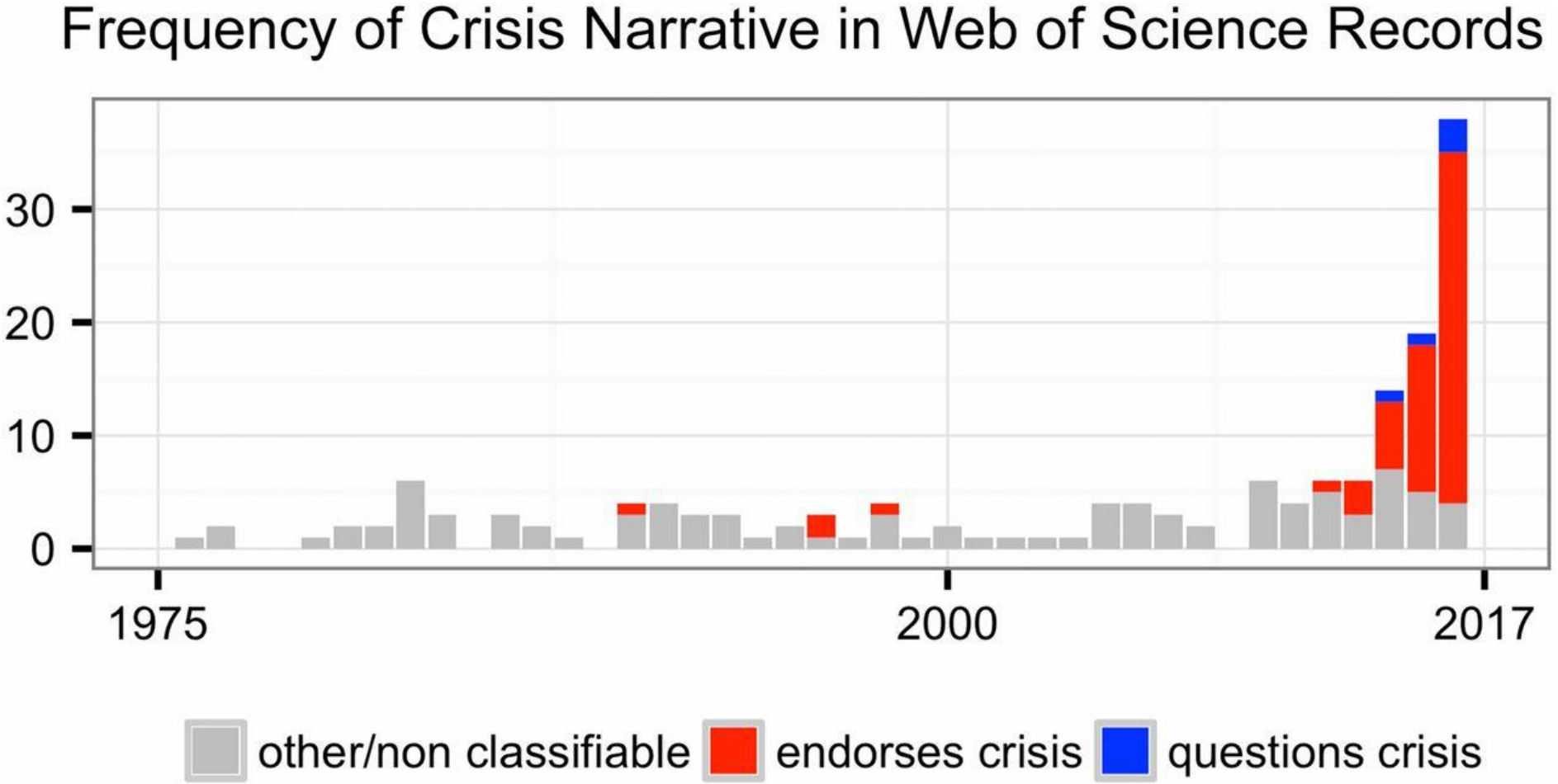
Replication and Reproducibility in Social Sciences and Statistics: Context, Concerns, and Concrete Measures

Lars Vilhuber

Cornell University

Partial funding acknowledged under NSF-#1131848 (NCRN) and a grant from the Alfred P. Sloan Foundation. The opinions expressed in this talk are solely the authors, and do not represent the views of the U.S. Census Bureau, the American Economic Association, or any of the funding agencies.

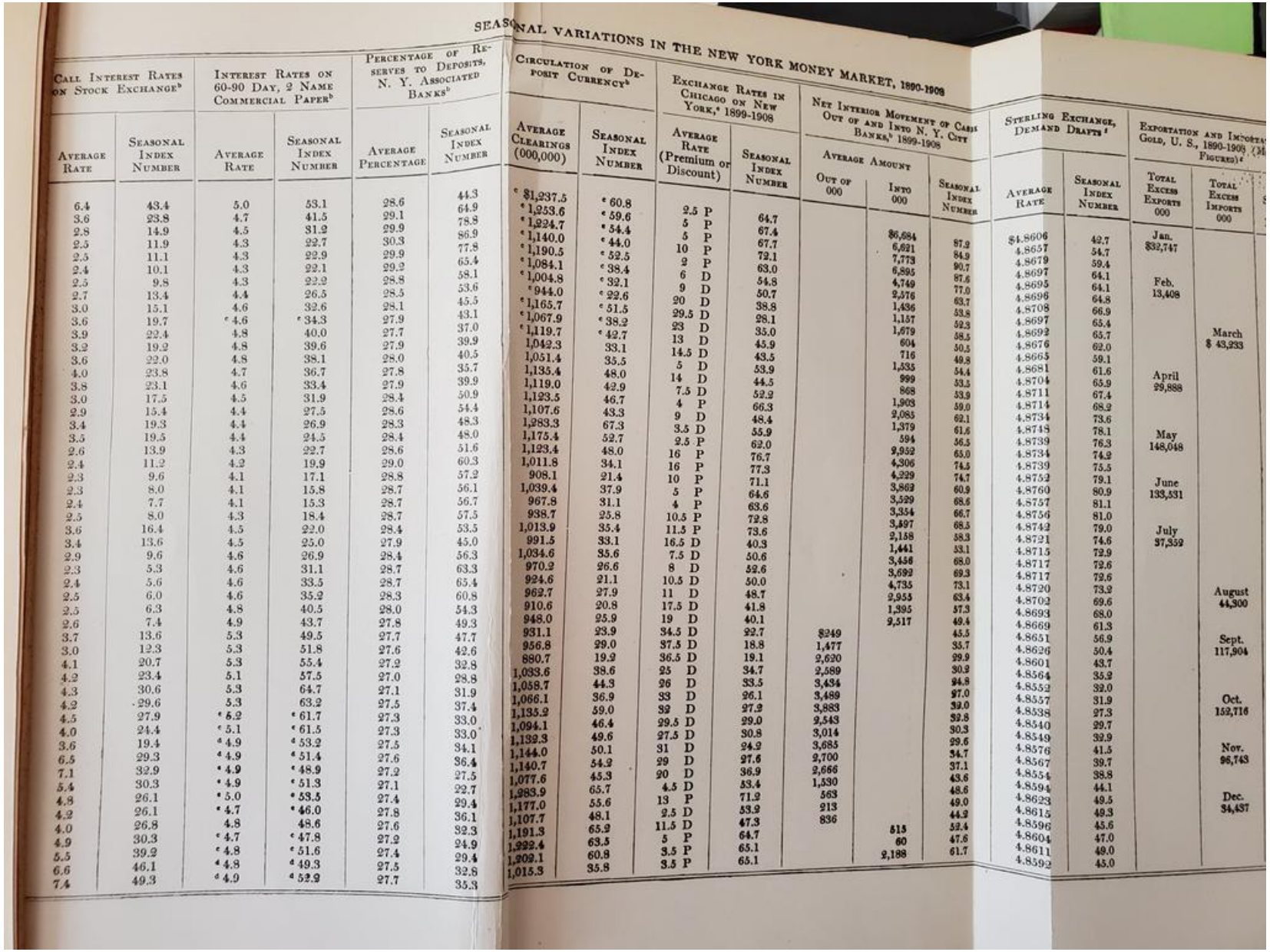
This reproducibility crisis thing….



https://doi.org/10.1073/pnas.1708272114

Efficiency of scholary discourse?

- Early publications (20th century) contained **tables of data**, and the **math** was simple (maybe)
  - **Data** became electronic, was no longer **included** or **cited**
  - **Math** was transcribed to **code**, and was no longer **included**



SEASONAL VARIATIONS IN THE NEW YORK MONEY MARKET, 1890-1908

Call Interest Rates on Stock Exchange[b]		Interest Rates on 60-90 Day, 2 Name Commercial Paper[b]		Percentage of Reserves to Deposits, N. Y. Associated Banks[b]		Circulation of Deposit Currency[b]		Exchange Rates in Chicago on New York,[c] 1899-1908		Net Interior Movement of Cash Out of and Into N. Y. City Banks,[b] 1899-1908			Sterling Exchange, Demand Drafts[d]		Exportation and Importation of Gold, U. S., 1890-1908 (Minus Figures)[e]	
Average Rate	Seasonal Index Number	Average Rate	Seasonal Index Number	Average Percentage	Seasonal Index Number	Average Clearings (000,000)	Seasonal Index Number	Average Rate (Premium or Discount)	Seasonal Index Number	Average Amount Out of 000	Average Amount Into 000	Seasonal Index Number	Average Rate	Seasonal Index Number	Total Excess Exports 000	Total Excess Imports 000
6.4	43.4	5.0	53.1	28.6	44.3	[c] $1,237.5	[c] 60.8	2.5 P	64.7		$6,684	87.2	$4.8606	42.7	Jan. $32,747	
3.6	23.8	4.7	41.5	29.1	64.9	[c] 1,253.6	[c] 59.6	5 P	67.4		6,621	84.9	4.8657	54.7		
2.8	14.9	4.5	31.2	29.9	78.8	[c] 1,224.7	[c] 54.4	5 P	67.7		7,773	90.7	4.8679	59.4		
2.5	11.9	4.3	22.7	30.3	86.9	[c] 1,140.0	[c] 44.0	10 P	72.1		6,895	87.6	4.8697	64.1	Feb. 13,408	
2.5	11.1	4.3	22.9	29.9	77.8	[c] 1,190.5	[c] 52.5	2 P	63.0		4,749	77.0	4.8695	64.1		
2.4	10.1	4.3	22.1	29.2	65.4	[c] 1,084.1	[c] 38.4	6 D	54.8		2,576	63.7	4.8696	64.8		
2.5	9.8	4.3	22.2	28.8	58.1	[c] 1,004.8	[c] 32.1	9 D	50.7		1,436	53.8	4.8708	66.9		
2.7	13.4	4.4	26.5	28.5	53.6	[c] 944.0	[c] 22.6	20 D	38.8		1,157	52.3	4.8697	65.4		
3.0	15.1	4.6	32.6	28.1	45.5	[c] 1,165.7	[c] 51.5	29.5 D	28.1		1,679	58.5	4.8692	65.7		March $ 43,233
3.6	19.7	[c] 4.6	[c] 34.3	27.9	43.1	[c] 1,067.9	[c] 38.2	23 D	35.0		604	50.5	4.8676	62.0		
3.9	22.4	4.8	40.0	27.7	37.0	[c] 1,119.7	[c] 42.7	13 D	45.9		716	49.8	4.8665	59.1		
3.2	19.2	4.8	39.6	27.9	39.9	1,042.3	33.1	14.5 D	43.5		1,535	54.4	4.8681	61.6		
3.6	22.0	4.8	38.1	28.0	40.5	1,051.4	35.5	5 D	53.9		999	53.5	4.8704	65.9	April 29,888	
4.0	23.8	4.7	36.7	27.8	35.7	1,135.4	48.0	14 D	44.5		868	53.9	4.8711	67.4		
3.8	23.1	4.6	33.4	27.9	39.9	1,119.0	42.9	7.5 D	52.2		1,903	59.0	4.8714	68.2		
3.0	17.5	4.5	31.9	28.4	50.9	1,123.5	46.7	4 P	66.3		2,085	62.1	4.8734	73.6		
2.9	15.4	4.4	27.5	28.6	54.4	1,107.6	43.3	9 D	48.4		1,379	61.6	4.8748	78.1	May 148,048	
3.4	19.3	4.4	26.9	28.3	48.3	1,283.3	67.3	3.5 D	55.9		594	56.5	4.8739	76.3		
3.5	19.5	4.4	24.5	28.4	48.0	1,175.4	52.7	2.5 P	62.0		2,952	65.0	4.8734	74.2		
2.6	13.9	4.3	22.7	28.6	51.6	1,123.4	48.0	16 P	76.7		4,306	74.5	4.8739	75.5		
2.4	11.2	4.2	19.9	29.0	60.3	1,011.8	34.1	16 P	77.3		4,229	74.7	4.8752	79.1	June 133,531	
2.3	9.6	4.1	17.1	28.8	57.2	908.1	21.4	10 P	71.1		3,863	69.9	4.8760	80.9		
2.3	8.0	4.1	15.8	28.7	56.1	1,039.4	37.9	5 P	64.6		3,529	68.6	4.8757	81.1		
2.4	7.7	4.1	15.3	28.7	56.7	967.8	31.1	4 P	63.6		3,354	66.7	4.8756	81.0		
2.5	8.0	4.3	18.4	28.7	57.5	938.7	25.8	10.5 P	72.8		3,497	68.5	4.8742	79.0	July 37,352	
3.6	16.4	4.5	22.0	28.4	53.5	1,013.9	35.4	11.5 P	73.6		2,158	58.3	4.8721	74.6		
3.4	13.6	4.5	25.0	27.9	45.0	991.5	33.1	16.5 D	40.3		1,441	53.1	4.8715	72.9		
2.9	9.6	4.6	26.9	28.4	56.3	1,034.6	35.6	7.5 D	50.6		3,456	68.0	4.8717	72.6		
2.3	5.3	4.6	31.1	28.7	63.3	970.2	26.6	8 D	52.6		3,692	69.3	4.8717	72.6		
2.4	5.6	4.6	33.5	28.7	65.4	924.6	21.1	10.5 D	50.0		4,735	73.1	4.8720	73.2		August 44,300
2.5	6.0	4.6	35.2	28.3	60.8	962.7	27.9	11 D	48.7		2,955	63.4	4.8702	69.6		
2.5	6.3	4.8	40.5	28.0	54.3	910.6	20.8	17.5 D	41.8		1,395	57.3	4.8693	68.0		
2.6	7.4	4.9	43.7	27.8	49.3	948.0	25.9	19 D	40.1		2,517	49.4	4.8669	61.3		
3.7	13.6	5.3	49.5	27.7	47.7	931.1	23.9	34.5 D	22.7	$249		45.5	4.8651	56.9		Sept. 117,904
3.0	12.3	5.3	51.8	27.6	42.6	956.8	29.0	37.5 D	18.8	1,477		35.7	4.8626	50.4		
4.1	20.7	5.3	55.4	27.2	32.8	880.7	19.2	36.5 D	19.1	2,620		29.9	4.8601	43.7		
4.2	23.4	5.1	57.5	27.0	28.8	1,033.6	38.6	25 D	34.7	2,589		30.2	4.8564	35.2		
4.3	30.6	5.3	64.7	27.1	31.9	1,058.7	44.3	26 D	33.5	3,434		24.8	4.8552	32.0		
4.2	29.6	5.3	63.2	27.5	37.4	1,066.1	36.9	33 D	26.1	3,489		27.0	4.8557	31.9		Oct. 152,716
4.5	27.9	[c] 5.2	[c] 61.7	27.3	33.0	1,135.2	59.0	32 D	27.2	3,883		32.0	4.8538	27.3		
4.0	24.4	[c] 5.1	[c] 61.5	27.3	33.0	1,094.1	46.4	29.5 D	29.0	2,543		32.8	4.8540	29.7		
3.6	19.4	[d] 4.9	[d] 53.2	27.5	34.1	1,132.3	49.6	27.5 D	30.8	3,014		30.3	4.8549	32.9		
6.5	29.3	[d] 4.9	[d] 51.4	27.6	36.4	1,144.0	50.1	31 D	24.2	3,685		29.6	4.8576	41.5		Nov. 96,743
7.1	32.9	[d] 4.9	[d] 48.9	27.2	27.5	1,140.7	54.2	29 D	27.6	2,700		34.7	4.8567	39.7		
5.4	30.3	[d] 4.9	[d] 51.3	27.1	22.7	1,077.6	45.3	20 D	36.9	2,666		37.1	4.8554	38.8		
4.8	26.1	[d] 5.0	[d] 53.5	27.4	29.4	1,283.9	65.7	4.5 D	53.4	1,530		43.6	4.8594	44.1		
4.2	26.1	[d] 4.7	[d] 46.0	27.8	36.1	1,177.0	55.6	13 P	71.2	563		48.6	4.8623	49.5		Dec. 34,437
4.0	26.8	4.8	48.6	27.6	32.3	1,107.7	48.1	2.5 D	53.2	213		49.0	4.8615	49.3		
4.9	30.3	[c] 4.7	[c] 47.8	27.2	24.9	1,191.3	65.2	11.5 D	47.3	836		44.2	4.8596	45.6		
5.5	39.2	[c] 4.8	[c] 51.6	27.4	29.4	1,222.4	63.5	5 P	64.7		513	52.4	4.8604	47.0		
6.6	46.1	[d] 4.8	[d] 49.3	27.5	32.8	1,202.1	60.8	3.5 P	65.1		60	47.6	4.8611	49.0		
7.4	49.3	[d] 4.9	[d] 52.2	27.7	35.3	1,015.3	35.8	3.5 P	65.1		2,188	61.7	4.8592	45.0		

From @sdellavi
AER 1911

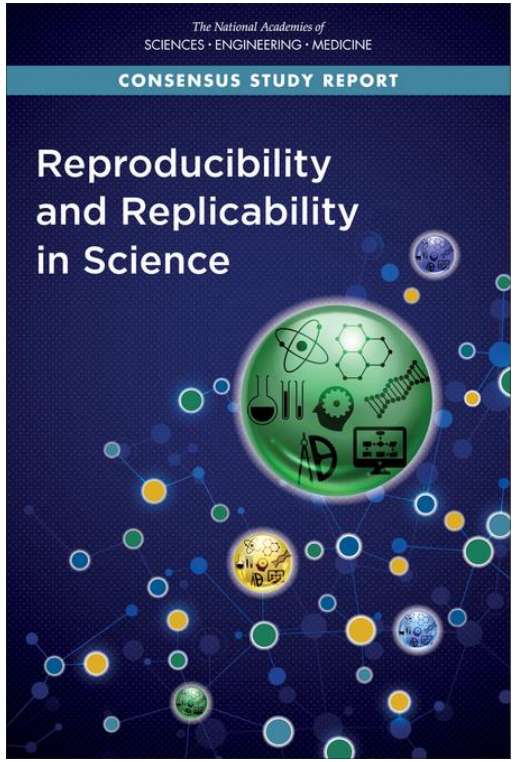
Efficiency of scholary discourse!

Modern publications thus need
the same transparency and completeness
as in the old days

to facilitate replicability

Replication continuum

https://doi.org/10.17226/25303

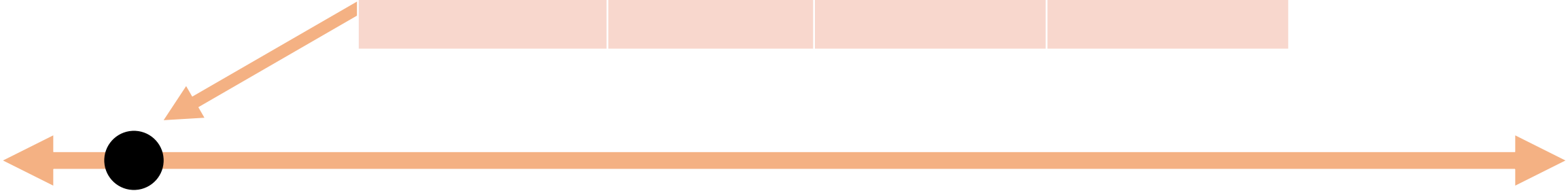


Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replication continuum

Same data	Same code	Same methods	Same context



Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replication continuum

Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Replication continuum

Same data	**Different code or software**	Same methods	Same context

Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Replication continuum

New data	Same code	Same methods	Same context
collection			

Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Replication continuum

Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Generalizability

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)

Replication continuum

Different data	Different code	**Different**	**Different**
	or software	**methods**	**context or**
			country

Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Generalizability

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)

Issues

Economics makes wide use of public-use data

- **Macrodata:**

"We use data downloaded from the Bureau of Economic Analysis…"

- **Microdata**:

"… this paper uses data from the Current Population Survey…"

This should be easy!

Examples

1 **Bureau of Labor Statistics.** 2000–2010. "Current Employment Statistics: Colorado, Total Nonfarm, Seasonally adjusted - SMS08000000000000001." United States Department of Labor. http://data.bls.gov/cgi-bin/surveymost?sm+08 (accessed February 9, 2011).

https://www.aeaweb.org/journals/policies/sample-references

Problems Making RELIABLE archives

Many datasets

- Are imperfectly described
  - Very few data citations
- Are badly documented
- Have no (permanent) location defined
  - Even for data from high-profile organizations!
- All of the above

Making USEFUL archives

- From analysis of code from 1996 to 2003 (MMH2006):

"Other authors seem to think that the entire world shares the exact same hard drive layout, with "C:\MYDATA\MYPROJECT\" **sprinkled liberally** throughout their code. Of course, a would-be replicator has to **find and change all these**."

"The author might not realize all the data/subroutine files that his code utilizes, and **forget to include** said data/subroutine in his replication files."

Risk

Google

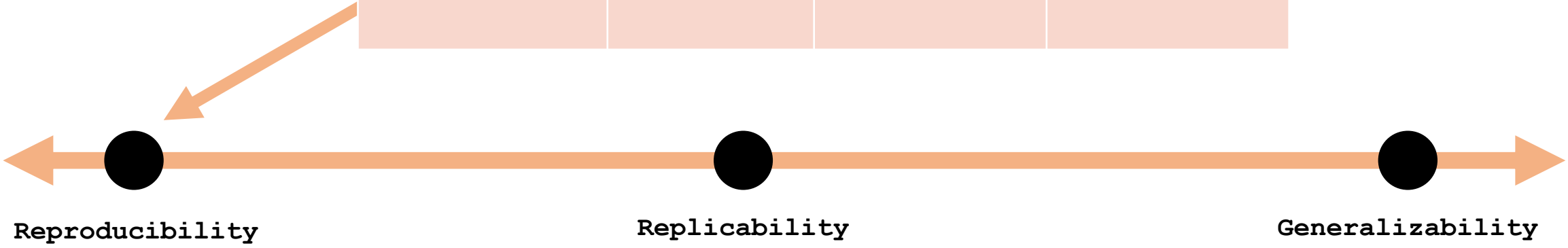
404. That's an error.

The requested URL /a_cool_website was not found on this server. That's all we know.

Not enough articles are reproducible

Replication continuum

Same data	Same code	Same methods	Same context



Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Generalizability

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)

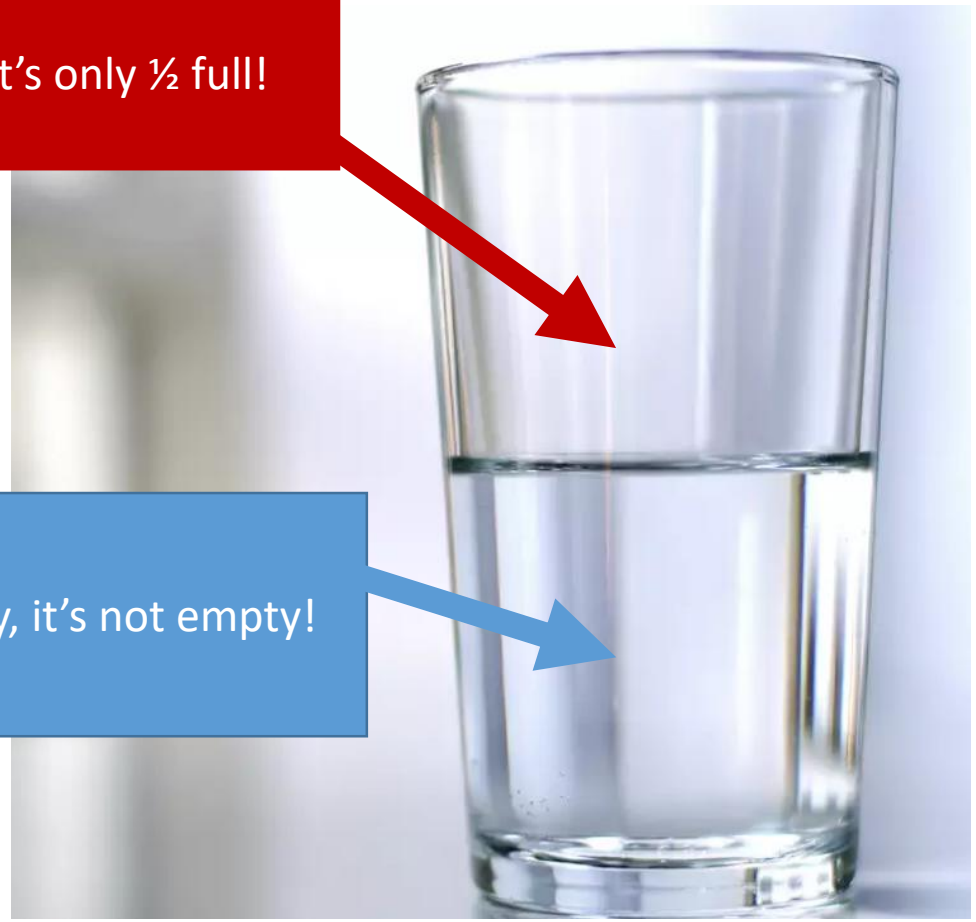
Some key statistics

Study	Year	N	Success	Type	Type-R	Type-Data	Percent	Field
Dewald Thursby Anderson	1986	54	2	Complete	Reproducibility	Avail	4%	Economics
Dewald Thursby Anderson	1986	54	7	Partial	Reproducibility	Avail	13%	Economics
McCullough McGeary Harrison	2006	186	14	Complete	Reproducibility	All	8%	Economics
McCullough McGeary Harrison	2006	62	14	Complete	Reproducibility	Avail	23%	Economics
Nosek et al	2015	100	36	Complete	Replication		36%	Psychology
Camerer et al	2016	18	11	Complete	Replication		61%	Experimental Econ
Chang Li	2017	67	29				33%	Macroeconomi
Kingi et al	2018	274	69	Complete	Reproducibility	All	25%	Economics
Kingi et al	2018	162	69	Complete	Reproducibility	Avail	43%	Economics
Kingi et al	2018	162	137	Partial	Reproducibility	Avail	85%	Economics

Kingi et al numbers are preliminary. Do not cite or quote.

In a nutshell

- **40%** use restricted-access data
- **25%** use public-use data and are mostly or completely reproducible
- **25%** use public-use data and are only partially reproducible
- **10%** fail to yield useful results



Current efforts at the AEA

- Pre-emptively improve code archives
  - By conducting reproducibility checks when we can
  - By working with groups that conduct reproducibility checks when we cannot

Current efforts at the AEA

- Better archives
  - Greater transparency of the code and data archives
  - Better provenance tracking
    - Leave code where it is when appropriate
    - Leave data where it is almost always
    - Display that information

AEA "Data Availability Policy" (2019)

- It is the policy of the American Economic Association to publish papers only if the data used in the analysis are **clearly and precisely** documented and **access to the data and code is clearly and precisely documented and is non-exclusive to the authors.**
- Authors of accepted papers that contain empirical work, simulations, or experimental work must **provide**, **prior to acceptance**, the data, programs, and other details of the computations **sufficient to permit replication**, as well as **information about access to data and programs.**

AEA "Data Availability Policy" (2018-> 2019)

- These will be **posted on the AEA website**. 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.

- Data and programs should **be archived in the AEA Data and Code Repository**. Authors will **provide access** to editors and reviewers, if requested**, to both data and programs prior to acceptance**. The Editor should be notified at the time of submission if access to the data used in a paper is restricted or limited, or if, for some other reason, the requirements above cannot be met. **The AEA Data Editor will assess compliance with this policy, and will verify the accuracy of the information prior to acceptance by the Editor.**

Action: Reproducibility Check

Social Science Data Editors

Data and Code Guidance by Data Editors

Guidance for authors wishing to create data and code supplements, and for replicators.

Verification guidance

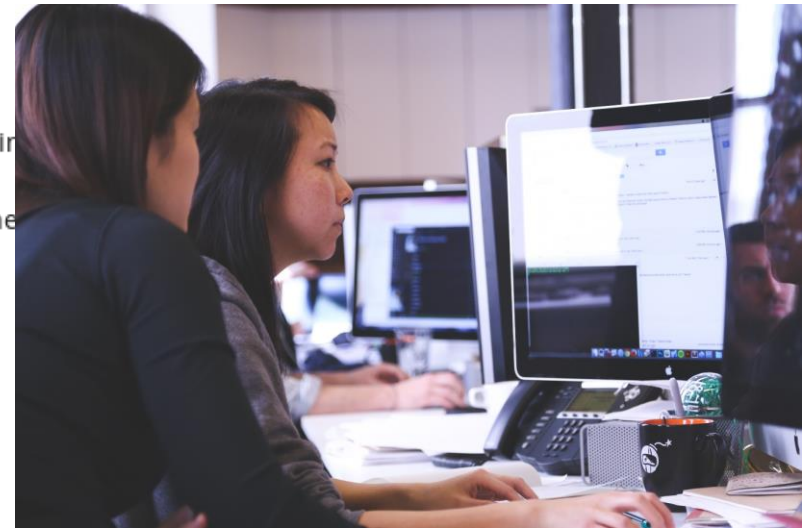
On this page:

- Overview
- Review the README file
- For each listed data source
- For each listed table, figure, in-text number
- Conduct a code verification, if data is available
- Examples

Overview

This document describes

- what authors should check before providir journals
- what verifier teams should check for in the to them for the purpose of verification



Goal: Improve reproducibility