

Reproducibility in research – Ensuring the transparency and credibility of your work

Lars Vilhuber Marie Connolly

2026-05-26



Follow along



larsvilhuber.github.io/summer-school-sfu-ubc-2026/presentation/presentation.html (PDF)

Who are we?



Your instructors



Marie Connolly



Lars Vilhuber

Marie Connolly

Marie Connolly is full professor at the Department of Economics, University of Quebec in Montreal's School of Management (ESG UQAM). Her research is primarily empirical and touches upon various topics in labor economics, such as intergenerational income mobility, the formation of human capital, the gender and family gap, women's labor force participation and the evaluation of public policy. She currently serves as the Data Editor of the Canadian Journal of Economics.



Lars Vilhuber

Executive Director of the **Labor Dynamics Institute** and Senior Research Associate in the **Economics Department** at **Cornell University**, and the **American Economic Association's** Data Editor.

American Economic Review



The *American Economic Review* is a general-interest economics journal. Established in 1911, the AER is among the nation's oldest and most respected scholarly journals in economics.

American Economic Review: Insights



AER: Insights is designed to be a top-tier, general-interest economics journal publishing papers of the same quality and importance as those in the AER, but devoted to publishing papers with important insights that can be conveyed succinctly.

Journal of Economic Literature



The *Journal of Economic Literature (JEL)*, first published in 1969, is designed to help economists keep abreast of and synthesize the vast flow of literature.

Journal of Economic Perspectives



The *Journal of Economic Perspectives (JEP)* fills the gap between the general interest press and academic economics journals.

American Economic Journal: Applied Economics



American Economic Journal: Applied Economics publishes papers covering a range of topics in applied economics, with a focus on empirical microeconomic issues.

American Economic Journal: Economic Policy



American Economic Journal: Economic Policy publishes papers covering a range of topics, the common theme being the role of economic policy in economic outcomes.

American Economic Journal: Macroeconomics



American Economic Journal: Macroeconomics focuses on studies of aggregate fluctuations and growth, and the role of policy in that context.

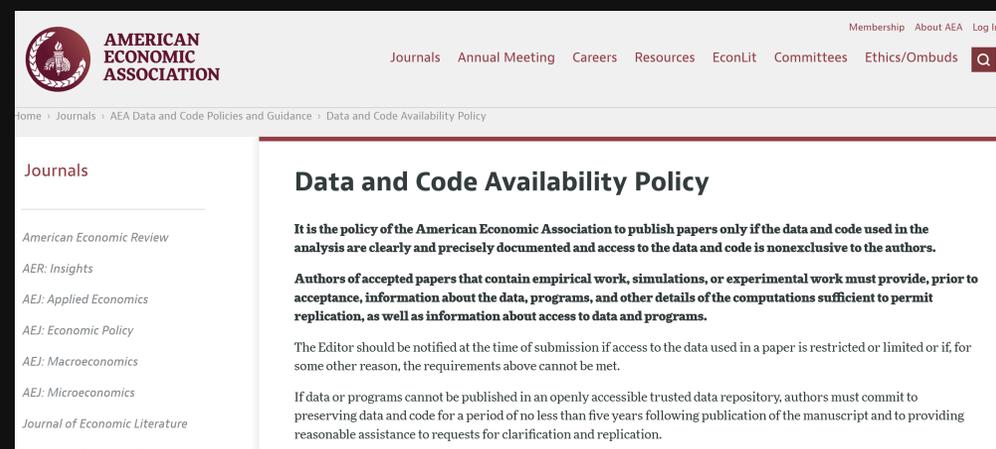
American Economic Journal: Microeconomics



American Economic Journal: Microeconomics publishes papers focusing on microeconomic theory; industrial organization; and the microeconomic aspects of international trade, political economy, and finance.

Data Editor of the AEA

2389 Manuscripts and 4440 Reports, approx. 4400 authors reached.



The screenshot shows the American Economic Association (AEA) website. The header includes the AEA logo and navigation links for Journals, Annual Meeting, Careers, Resources, EconLit, Committees, and Ethics/Ombuds. The main content area is titled "Data and Code Availability Policy" and contains the following text:

It is the policy of the American Economic Association to publish papers only if the data and code used in the analysis are clearly and precisely documented and access to the data and code is nonexclusive to the authors.

Authors of accepted papers that contain empirical work, simulations, or experimental work must provide, prior to acceptance, information about the data, programs, and other details of the computations sufficient to permit replication, as well as information about access to data and programs.

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.

If data or programs cannot be published in an openly accessible trusted data repository, authors must commit to preserving data and code for a period of no less than five years following publication of the manuscript and to providing reasonable assistance to requests for clarification and replication.

DCAP



Some housekeeping



Languages

- All slides are in **English**
- Primary language is **English**
- Toutes les questions peuvent être en ***français***
- Toutes les réponses vont être dans la ***langue posée***,
ou en anglais

Code of Conduct

- see [full Code of Conduct of LDI Replication Lab, Canadian Economics Association \(in French\)](#)

We are dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity. By participating in this team, participants accept to abide by LDI ReplicationLab's Code of Conduct and accept the procedures by which any Code of Conduct incidents are resolved. Any form of behaviour to exclude, intimidate, or cause discomfort is a violation of the Code of Conduct. In order to foster a positive and professional learning environment we encourage the following kinds of behaviours in all platforms and events:

Code of Conduct

- Use **welcoming** and inclusive language
- Be **respectful** of different viewpoints and experiences
- Gracefully accept **constructive** criticism
- Show courtesy and respect towards other community members

If you believe someone is violating the Code of Conduct, we ask that you report it to us, QICSS/CIQSS (sponsor), or the CEA (host).



Walkthrough of the agenda



Today

- 9:00 Welcome
- 9:05 Walkthrough
- 9:15 Goals
- 9:30 Technical setup, possible team formation
- 9:45  *Hands-on Exercise: A very imperfect example*
- 10:00 Discussion of the “Very imperfect example”
- 10:30 Break (15 minutes)
- 10:45 **Day 1:**
 - Setting yourself up for reproducibility

- 12:00 Lunch Break (30 minutes)



- 12:30 **How to run Stata!** or R! (reproducibly)
- 12:45 Extra: **How to install Stata packages**
- 1:00 **Topic A** (see Survey)
- 1:45 **Topic B** (see Survey)
- 2:30 **Hands-on:** Improving the replication package (very imperfect -> a lot better)
- 2:50 **Hands-on:** Testing it all
- 2:55 Wrap up

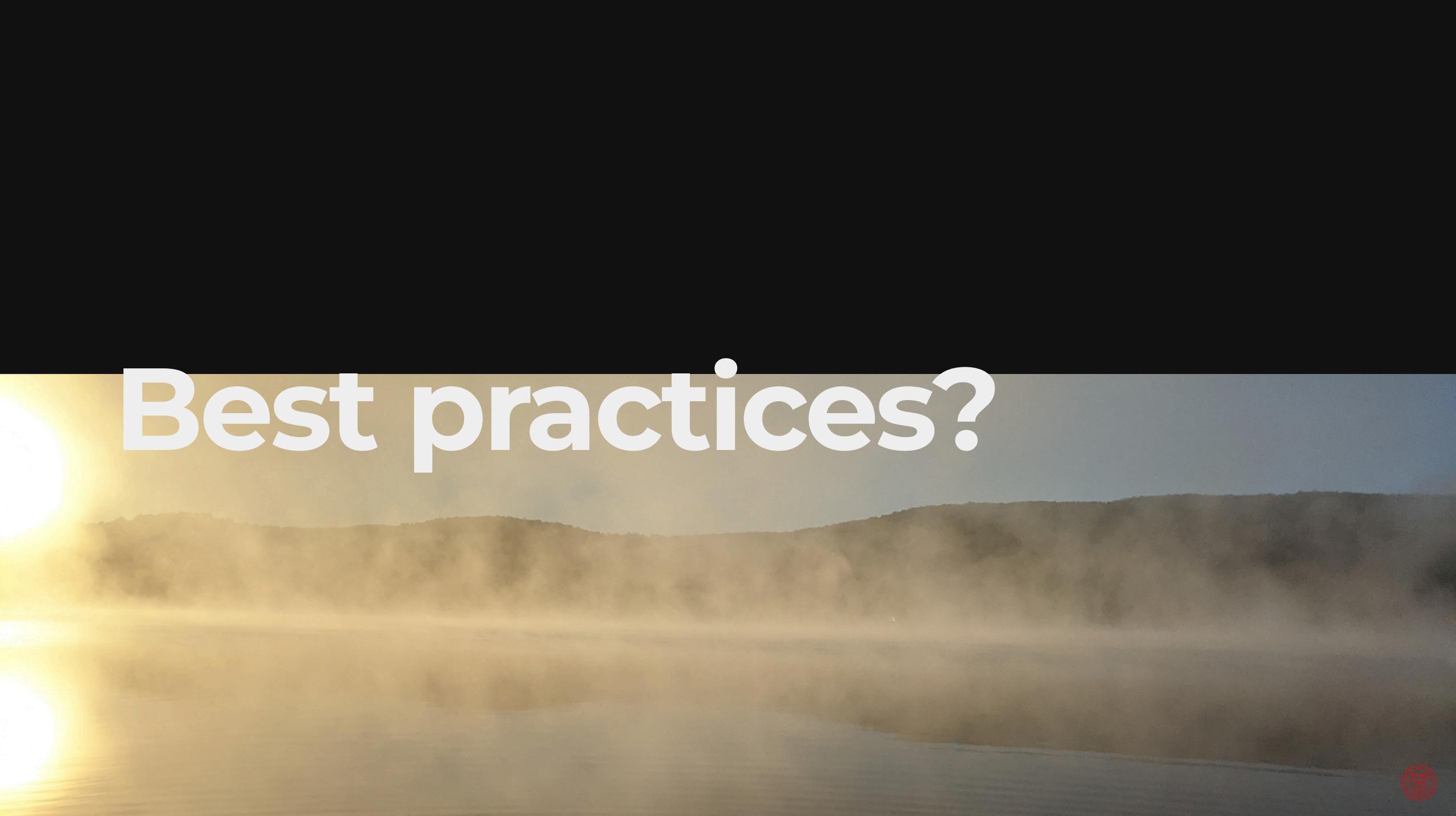
Tomorrow

- 9:00 Welcome
- 9:05 Walkthrough
- 9:15 Goals
- 9:30 Technical setup, possible team formation
- 9:45  *Hands-on Exercise: A very imperfect example*
- 10:00 Discussion of the “Very imperfect example”
- 10:30 Break (15 minutes)
- 10:45 **Day 1:**
 - Setting yourself up for reproducibility

- 12:00 Lunch Break (30 minutes)

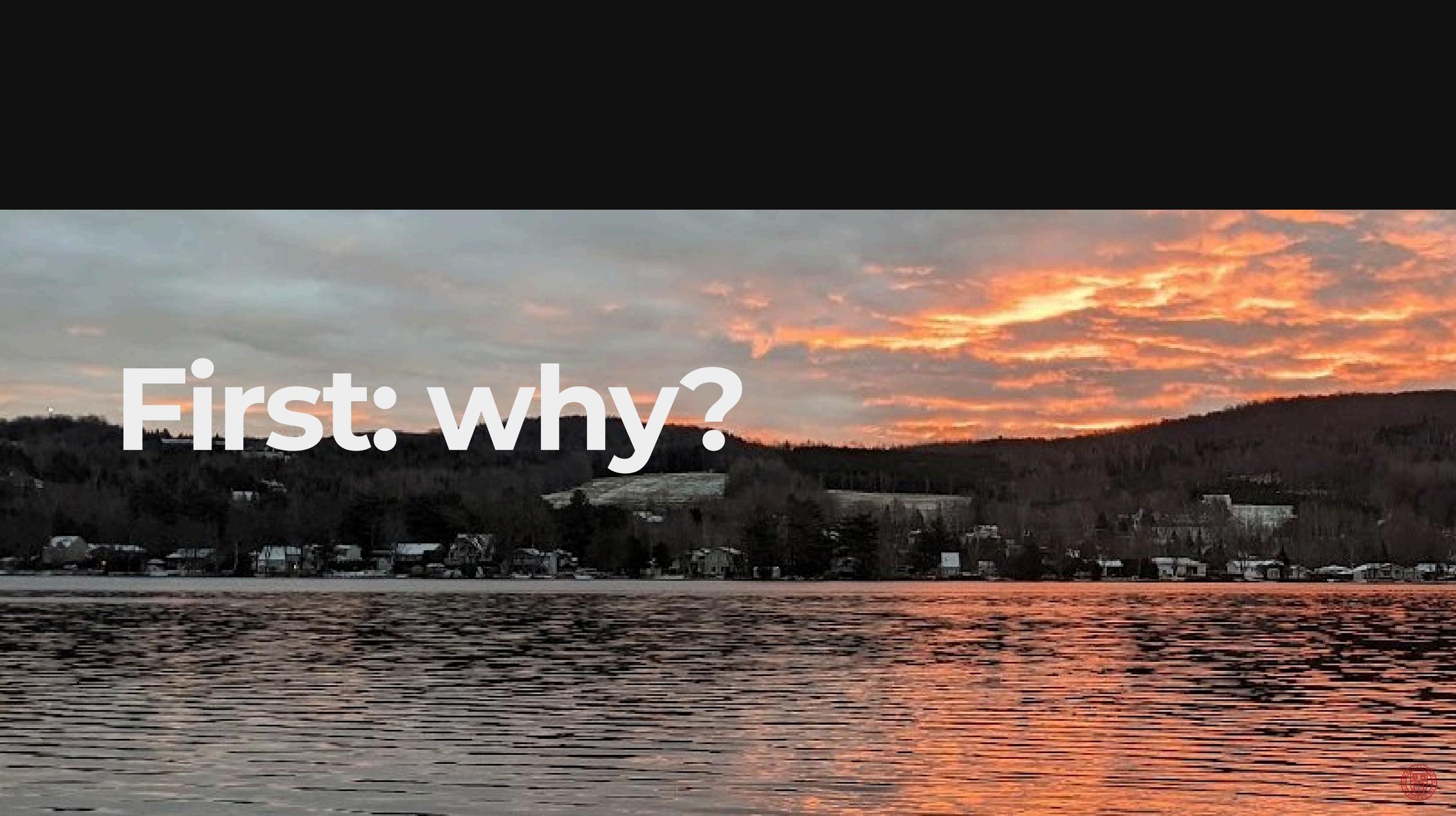


- 12:30 **How to run Stata!** or R! (reproducibly)
- 12:45 Extra: **How to install Stata packages**
- 1:00 **Topic A** (see Survey)
- 1:45 **Topic B** (see Survey)
- 2:30 **Hands-on:** Improving the replication package (very imperfect -> a lot better)
- 2:50 **Hands-on:** Testing it all
- 2:55 Wrap up



Best practices?





First: why?



Why reproducibility?

- Credibility
- Transparency (openness)
- Efficiency of scholarly discourse?

Why reproducibility?

- Early publications (20th century) **contained tables of data**, and the math was simple (maybe)
- Data became electronic, **were 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

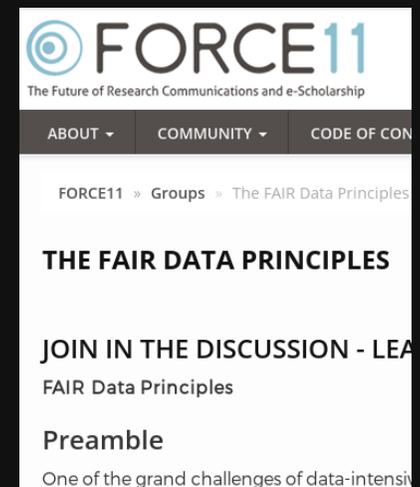
CALL INTEREST RATES ON STOCK EXCHANGE*				INTEREST RATES ON 60-90 DAY, 9 NAME COMMERCIAL PAPER*				PERCENTAGE OF RESERVES TO DEPOSITORS, N. Y. ASSOCIATED BANKS*				CIRCULATION OF DEPOSIT CURRENCY*				EXCHANGE RATES IN CHICAGO ON NEW YORK, 1890-1908				NEW INTERIOR MOVEMENT OF CASH OUT OF AND INTO N. Y. CITY BANKS, 1890-1908				STERLING EXCHANGE DEMAND DRAHTS*				EXPORTATION AND IMPORTATION GOLD, U. S., 1890-1908 (24 MONTHS)			
AVERAGE RATE	SEASONAL INDEX NUMBER	AVERAGE RATE	SEASONAL INDEX NUMBER	AVERAGE PERCENTAGE	SEASONAL INDEX NUMBER	AVERAGE PERCENTAGE	SEASONAL INDEX NUMBER	AVERAGE CIRCULATION (000,000)	SEASONAL INDEX NUMBER	AVERAGE RATE (PERCENT OF DISCOUNT)	SEASONAL INDEX NUMBER	AVERAGE AMOUNT OUT OF 000	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			* 81,327.5	* 60.8	2.5 P	64.7																				
5.5	23.8	4.7	41.1	29.1	64.9			* 1,233.6	* 69.6	5 P	67.4																				
2.3	14.9	4.5	31.3	29.9	78.8			* 1,224.7	* 44.4	5 P	67.4																				
2.5	11.9	4.3	22.7	30.3	80.9			* 1,140.0	* 44.0	10 P	67.7																				
2.1	11.1	4.3	22.9	29.9	77.8			* 1,190.5	* 42.3	3 P	72.1																				
2.4	10.1	4.3	22.1	29.2	65.4			* 1,084.1	* 38.4	6 D	63.0																				
2.5	9.8	4.3	22.2	28.8	58.1			* 1,048.3	* 32.1	6 D	54.5																				
2.7	13.4	4.4	26.5	28.3	53.6			* 944.0	* 22.2	9 D	50.7																				
3.0	13.1	4.6	32.6	28.1	45.5			* 1,165.7	* 51.5	20 D	38.8																				
3.6	19.7	4.6	* 34.3	27.9	43.1			* 1,067.9	* 38.2	13 D	28.1																				
3.9	22.4	4.8	40.0	27.7	37.0			* 1,119.7	* 42.7	23 D	33.0																				
3.2	19.2	4.8	39.6	27.9	39.9			1,042.3	33.1	13 D	43.9																				
3.6	22.0	4.8	38.1	28.0	30.9			1,031.4	35.5	15 D	43.5																				
4.0	23.8	4.7	36.7	27.8	35.7			1,133.4	48.0	5 D	53.9																				
3.8	23.1	4.6	33.4	27.9	38.9			1,119.0	42.9	14 D	44.5																				
3.0	17.5	4.5	31.9	28.4	40.9			1,123.5	46.7	9 D	52.2																				
2.9	15.4	4.4	27.5	28.6	54.4			1,107.6	43.3	4 P	66.3																				
3.4	19.3	4.4	26.9	28.3	48.3			1,282.3	67.3	3.5 D	55.9																				
3.5	19.5	4.4	24.5	28.4	48.0			1,172.4	52.7	2.5 P	62.0																				
2.6	13.9	4.3	22.7	28.6	51.6			1,123.4	48.0	16 P	76.7																				
2.4	11.2	4.2	18.9	29.0	60.3			1,011.8	34.1	16 P	77.3																				
2.3	9.6	4.1	17.1	28.8	57.2			928.1	21.4	10 P	71.1																				
2.3	8.0	4.1	15.8	28.7	56.1			1,039.4	37.9	5 P	64.1																				
2.4	7.7	4.1	13.3	28.7	56.7			997.9	31.1	4 P	68.6																				
2.5	8.0	4.3	18.4	28.7	37.5			938.7	35.8	10.5 P	72.8																				
3.6	16.4	4.5	22.0	28.4	33.5			1,013.9	33.4	11.5 P	73.6																				
3.4	13.6	4.5	25.0	27.9	45.0			891.3	33.1	16.5 D	40.3																				
2.9	9.6	4.6	20.9	28.4	36.3			1,054.6	33.6	7.5 D	50.6																				
2.3	3.3	4.6	31.1	28.7	63.3			970.2	26.6	8 D	52.6																				
2.4	5.6	4.6	33.3	28.7	63.4			894.6	21.1	10.5 D	50.0																				
2.5	6.0	4.6	33.3	28.3	60.8			862.7	27.9	11 D	48.7																				
2.5	6.3	4.8	40.5	28.0	54.3			810.6	20.8	17.5 D	41.8																				
2.6	7.4	4.9	43.7	27.8	49.3			848.0	23.9	19 D	40.1																				
3.7	13.6	5.3	49.5	27.7	47.7			831.1	23.9	34.5 D	32.7																				
3.0	12.3	5.3	51.8	27.6	42.6			866.8	29.0	37.5 D	18.8																				
4.1	20.7	5.3	55.4	27.2	32.8			880.7	19.2	36.5 D	18.1																				
4.2	23.4	5.1	37.5	27.0	28.8			1,038.6	38.6	25 D	34.7																				
4.3	30.6	5.3	64.7	27.1	31.9			1,038.7	44.3	26 D	33.3																				
4.2	29.6	5.3	63.2	27.5	37.4			1,066.1	36.9	33 D	29.1																				
4.5	37.9	* 4.2	* 61.7	27.3	33.0			1,135.2	59.0	33 D	27.3																				
4.0	24.4	* 5.1	* 61.5	27.3	33.0			1,094.1	46.4	29.5 D	29.0																				
3.6	19.4	* 4.9	* 53.2	27.5	34.1			1,140.7	54.2	29 D	27.6																				
6.5	29.3	* 4.9	* 51.4	27.6	36.4			1,077.6	43.3	30 D	26.9																				
7.1	32.9	* 4.9	* 48.9	27.2	27.5			1,065.9	65.7	4.5 D	33.4																				
5.4	30.3	* 4.9	* 51.3	27.1	22.7			1,177.0	55.4	13 P	21.8																				
4.8	26.1	* 5.0	* 55.5	27.4	36.1			1,077.7	48.1	2.5 D	33.9																				
4.2	26.1	* 4.7	* 46.0	27.4	30.4			1,101.3	65.2	11.5 D	47.3																				
4.0	26.8	4.8	48.6	27.6	32.3			1,222.4	63.2	5 P	44.7																				
4.9	30.3	* 4.7	* 47.8	27.2	34.9			1,202.1	60.8	3.5 P	65.1																				
6.5	39.2	* 4.8	* 51.6	27.4	29.4			1,015.3	35.8	3.5 P	65.1																				
6.6	46.1	* 4.8	* 49.3	27.5	32.8																										
7.4	49.3	* 4.3	* 52.2	27.7	35.3																										

AER 1911



Increasing broad consensus in academia

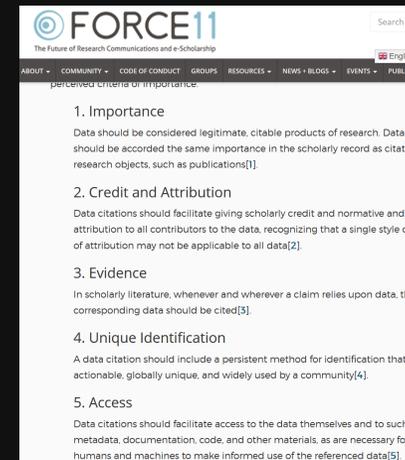
- **FAIR principles**
- Data Citation Principles
- Computational Reproducibility
- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable



Data Citation Principles

- FAIR principles
- **Data Citation Principles**
- Computational Reproducibility

- To make it **findable,**
citations,
- Give **attribution**
and **credit** for
data.



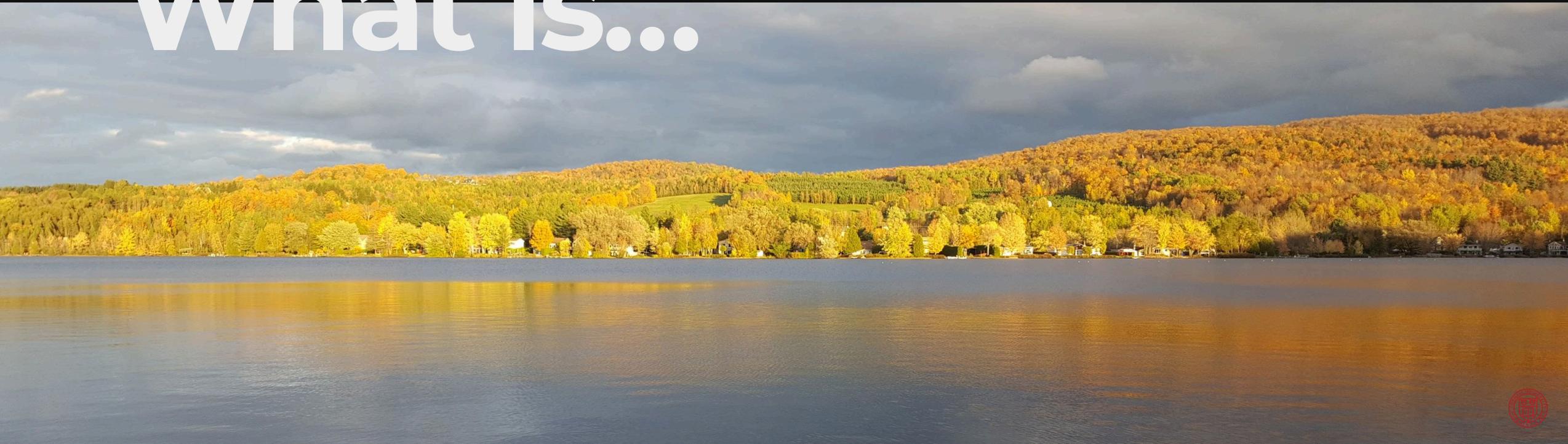
Computational reproducibility

- FAIR principles
- Data Citation Principles
- **Computational Reproducibility**

- Primary topic today

Reproducibility means obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis.²

What is...



What is a replication package?

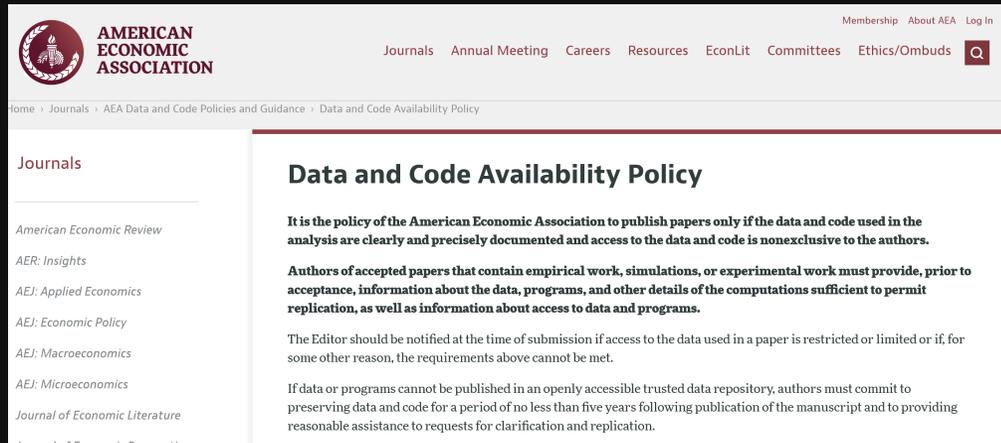


A Replication Package is

- Code
- Data
- Materials (for surveys, experiments, ...)
- Instructions on how to obtain data not included
- Instructions on how to combine it all
- Known issues documented

Complies with...

- AEA Data and Code Availability policy
- Data and Code Availability Standard 



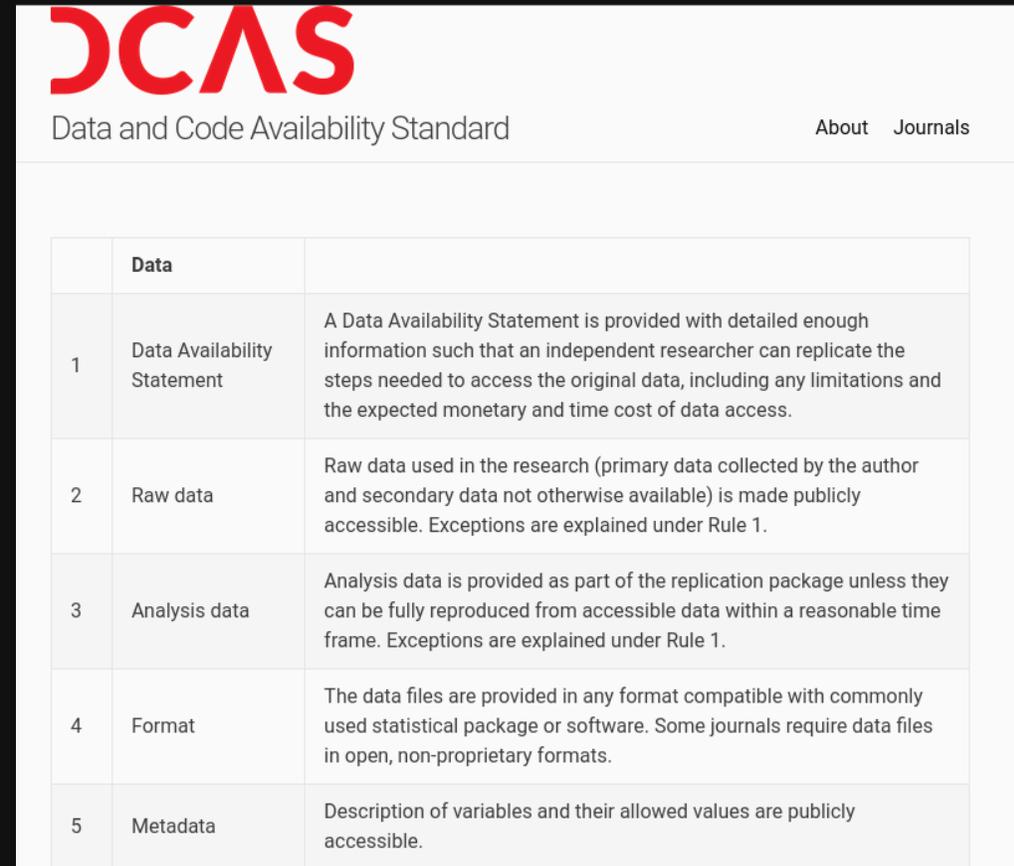
The screenshot shows the American Economic Association website. The header includes the AEA logo and navigation links for Journals, Annual Meeting, Careers, Resources, EconLit, Committees, Ethics/Ombuds, and a search icon. The main content area is titled "Data and Code Availability Policy" and contains the following text:

It is the policy of the American Economic Association to publish papers only if the data and code used in the analysis are clearly and precisely documented and access to the data and code is nonexclusive to the authors.

Authors of accepted papers that contain empirical work, simulations, or experimental work must provide, prior to acceptance, information about the data, programs, and other details of the computations sufficient to permit replication, as well as information about access to data and programs.

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.

If data or programs cannot be published in an openly accessible trusted data repository, authors must commit to preserving data and code for a period of no less than five years following publication of the manuscript and to providing reasonable assistance to requests for clarification and replication.



The screenshot shows the DCAS Data and Code Availability Standard website. The header includes the DCAS logo and navigation links for About and Journals. The main content area is titled "Data and Code Availability Standard" and contains a table with the following data:

	Data	
1	Data Availability Statement	A Data Availability Statement is provided with detailed enough information such that an independent researcher can replicate the steps needed to access the original data, including any limitations and the expected monetary and time cost of data access.
2	Raw data	Raw data used in the research (primary data collected by the author and secondary data not otherwise available) is made publicly accessible. Exceptions are explained under Rule 1.
3	Analysis data	Analysis data is provided as part of the replication package unless they can be fully reproduced from accessible data within a reasonable time frame. Exceptions are explained under Rule 1.
4	Format	The data files are provided in any format compatible with commonly used statistical package or software. Some journals require data files in open, non-proprietary formats.
5	Metadata	Description of variables and their allowed values are publicly accessible.

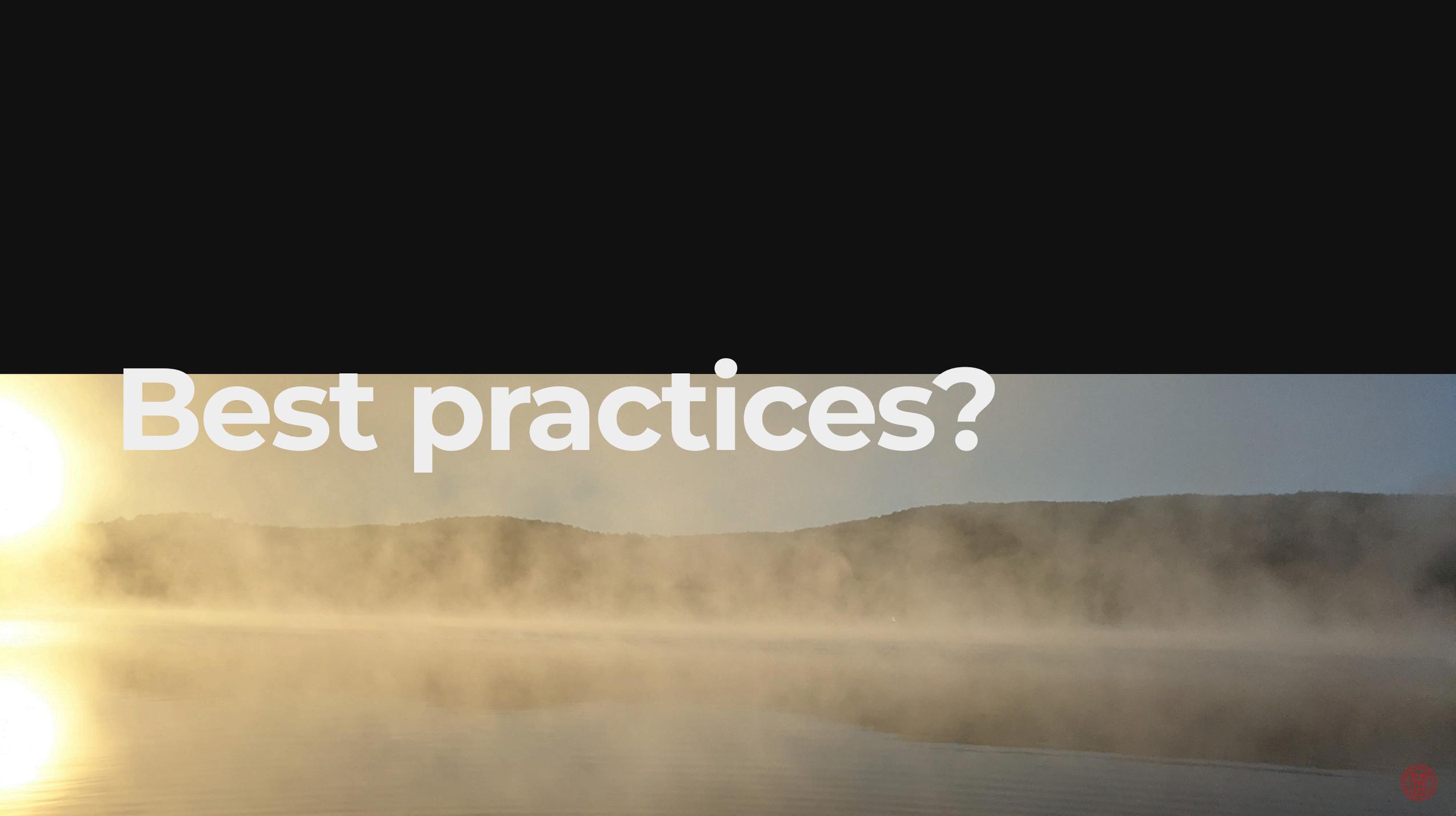


Is stored in...

- AEA Data and Code Repository
- Other trusted repositories

The screenshot shows the OPENICPSR interface for a project titled "Indirect Savings from Public Procurement Centralization". The page includes a header with the project name and navigation options like "Find Data", "Share Data", and "Repositories". Below the header, there is a section for "Principal Investigator(s)" listing Clarissa Lotti, Lear Arieda Muço, Giancarlo Spagnolo, and Tommaso Valletti. A "Version" dropdown is set to "v1". A table lists files and folders with columns for Name, File Type, Size, and Last Modified. The table includes folders for "code", "data", and "output", and files for "CITATION.CFF", "LICENSE.txt", "README.md", and "main.sh". On the right side, there is a "DOWNLOAD THIS FOLDER" button and a "Usage Metrics" section showing overall project metrics (14 Views, 3 Downloads, 3 Publications) and folder/file-level metrics (0 Views, 0 Downloads).

Name	File Type	Size	Last Modified
code			06/18/2024 01:15:PM
data			06/18/2024 01:16:PM
output			06/18/2024 01:14:PM
CITATION.CFF	text/plain	862 bytes	06/18/2024 09:14:AM
LICENSE.txt	text/plain	1.2 KB	06/18/2024 09:14:AM
README.md	text/x-web-markdown	6 KB	06/18/2024 09:14:AM
main.sh	application/x-sh	2.4 KB	06/18/2024 09:14:AM



Best practices?



Summing up

- **Why**
 - Credibility
 - Transparency (openness)
 - Efficiency of scholarly discourse ([example])
- **How**
 - FAIR principles
 - Data Citation Principles
 - Computational Reproducibility
- As **Replication Packages**
 - Code
 - Data
 - Materials (for surveys, experiments, ...)
 - Instructions on how to obtain data not included
 - Instructions on how to combine it all
 - Known issues documented

Who?



Who?

-  Authors at **conditional acceptance**
-  Authors at **submission**
-  Authors at **beginning** of project
-  Experienced researchers
-  Junior researchers
-  Ph.D. students
-  Undergraduates

Who?

You.



You

 Now:

- more efficient development
- more efficient collaboration
- more assurance that “everything just works”

 Soon

- more efficient development across projects
- more efficient response to editors and referees
- ... while you are in a **new** institution, on a **new computer**, with three courses to **prep**, and

 OPEN ACCESS  PEER-REVIEWED

RESEARCH ARTICLE

Experience of irreproducibility as a risk factor for poor mental health in biomedical science doctoral students: A survey and interview-based study

Nasser Lubega, Abigail Anderson, Nicole C. Nelson 

Published: November 15, 2023 • <https://doi.org/10.1371/journal.pone.0293584>

Article

Authors

Metrics

Comments

Media Coverage

Peer Review



How?



How to create reproducible research?



Habits

- Reproducibility from Day 1
- Adopt reproducible habits
- Take notes when you do things, not after
- Use version control



Strategy

Computational empathy: think of the next person to run this - It could be **you** in 5 years!



Hands-on: A very imperfect example

Presentation of the example



Day 1 reproducibility



Day 1: How to not to organize your work

A tale of ineffective technical collaboration



Day 1: Setting yourself up for reproducibility

An approach to be reproducible from Day 1



End of Day 1



Appendix



Where to?



Choices

- Issue No. 1: **Data Access** and preservation ✨ [a]
- Issue No. 2: **Confidential data** - same ⬆️! [a] [b]
- Lifecycle checking: **Self-checking reproducibility** and **presentation**
- New challenges: **AI and Big Data** but same ⬆️!
- New methods: **Transparency outsourced or certified**
- Implementation in academia: **Students!**



Resources



README

Lars Vilhuber, Connolly, M., Koren, M., Llull, J., & Morrow, P. (2022). A template README for social science replication packages (v1.1). Social Science Data Editors.

<https://doi.org/10.5281/zenodo.7293838>

You can download the Word, LaTeX, or Markdown version of the README with lots of examples.



Other guidance

- Presentation on “Self Checking Reproducibility” and its associated [website](#)
- Guidance when (some) data are confidential: <https://labordynamicsinstitute.github.io/reproducibility-confidential/>
- Guidance for citations: <https://social-science-data-editors.github.io/guidance/addtl-data-citation-guidance.html>

Extra info

- This document's source:
<https://github.com/larsvilhuber/summer-school-sfu-ubc-2026>
- Licensed under  CC BY-NC 4.0

Sources

- Images: [NYT](#), [Bluesky 1](#), [Bluesky 2](#), [Ike Hayman/Wikimedia](#)

Details on Transparency, etc.



Transparency

- Provenance of the **data**
- Processing of the data, from raw data to results (code)

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.

Completeness

- All data needs to be identified and and access described
- All code needs to be described and provided
- All materials must be provided (survey forms, etc.)

Authors ... must provide, prior to acceptance, the **data, programs, and other details** of the computations **sufficient** to permit replication

Preservation

- All **data** needs to be preserved for future replicators
 - Ideally, within the replication package, subject to ToU, for convenience
 - Otherwise, in a **trusted repository**

Preservation

- **Code** must be in a trusted repository
 - Usually, within the replication package
 - Websites, Github, are **not acceptable**

Historically

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

CALL INTEREST RATES ON STOCK EXCHANGE ^a		INTEREST RATES ON 60-90 DAY, 2 NAME COMMERCIAL PAPER ^b		PERCENTAGE OF RESERVES TO DEPOSITS, N. Y. ASSOCIATED BANKS ^c		CIRCULATION OF DEPOSIT CURRENCY ^d		EXCHANGE RATES IN CHICAGO ON NEW YORK, 1899-1908		NET INTERIOR MOVEMENT OF CASH OUT OF AND INTO N. Y. CITY BANKS, 1899-1908			STERLING EXCHANGE, DEMAND DRAFTS ^e		EXPORTATION AND IMPORTATION OF GOLD, U. S., 1890-1909 (Millions of Dollars) ^f	
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	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	* \$1,937.5	* 60.8									
3.6	23.8	4.7	41.5	29.1	64.9	* 1,933.6	* 59.6	2.5 P	64.7							
2.8	14.9	4.5	31.2	29.9	78.8	* 1,924.7	* 54.4	5 P	67.4	86,684	6,621	\$4,8606	42.7	Jan.		
2.5	11.9	4.3	22.7	30.3	86.9	* 1,140.0	* 44.0	10 P	67.7	7,773	84.9	4,8657	54.7			
2.5	11.1	4.3	22.9	29.9	77.8	* 1,190.5	* 52.5	9 P	72.1	6,895	90.7	4,8679	59.4			
2.4	10.1	4.3	22.1	29.2	65.4	* 1,084.1	* 38.4	6 D	63.0	4,749	87.6	4,8697	64.1	Feb.		
2.5	9.8	4.3	22.2	28.8	58.1	* 1,044.8	* 32.1	9 D	54.8	2,276	77.9	4,8695	64.1	13,408		
2.7	13.4	4.4	26.5	28.5	53.6	* 1,444.0	* 22.6	20 D	50.7	1,436	63.7	4,8696	64.8			
3.0	15.1	4.6	32.6	29.1	45.5	* 1,165.7	* 51.5	29.5 D	38.8	1,137	53.5	4,8708	66.9			
3.6	19.7	* 4.6	* 34.3	27.9	43.1	* 1,067.9	* 38.2	23 D	28.1	1,679	52.3	4,8697	65.4			
3.9	22.4	4.8	40.0	27.7	37.0	* 1,119.7	* 42.7	13 D	35.0	604	38.5	4,8692	65.7			
3.2	19.2	4.8	39.6	27.9	39.9	1,042.3	33.1	14.5 D	43.5	716	30.5	4,8676	62.0		March \$ 43,293	
3.6	22.0	4.8	38.1	28.0	40.5	1,031.4	35.5	5 D	33.9	1,335	49.5	4,8665	59.1			
4.0	23.8	4.7	36.7	27.9	35.7	1,135.4	48.0	14 D	44.5	999	54.4	4,8681	61.6			
3.8	23.1	4.6	33.4	27.9	35.7	1,119.0	42.9	7.5 D	52.2	868	33.3	4,8704	65.9	April	29,888	
3.0	17.5	4.5	31.9	28.4	39.9	1,107.6	46.7	4 P	66.3	1,903	59.0	4,8714	68.2			
2.9	15.4	4.4	27.5	28.6	54.4	1,193.5	43.3	9 D	48.4	2,085	62.1	4,8734	73.6			
3.4	19.3	4.4	26.9	28.3	48.3	1,107.6	46.7	3.5 D	55.9	1,379	61.6	4,8748	78.1	May		
3.5	19.5	4.4	24.5	28.4	48.0	1,175.4	52.7	2.5 P	62.0	594	36.5	4,8739	76.3	148,048		
2.6	13.9	4.3	22.7	28.6	51.6	1,193.4	48.0	16 P	76.7	2,822	65.0	4,8734	74.2			
2.4	11.2	4.2	19.9	29.0	60.3	1,011.8	34.1	10 P	77.3	4,329	74.5	4,8739	75.5			
2.3	9.6	4.1	17.1	28.8	57.2	908.1	21.4	5 P	71.1	3,869	74.7	4,8732	79.1	June		
2.3	8.0	4.1	15.8	28.7	56.1	1,039.4	37.9	4 P	64.6	3,569	68.6	4,8760	80.9	133,531		
2.4	7.7	4.1	15.3	28.7	56.7	967.8	31.1	10.5 P	63.6	3,529	66.7	4,8757	81.1			
2.5	8.0	4.3	18.4	28.7	57.5	938.7	25.8	8 D	52.6	3,354	68.7	4,8756	81.0			
3.6	16.4	4.5	22.0	28.4	53.5	1,013.9	35.4	11.5 P	72.8	3,497	68.5	4,8742	79.0	July		
3.4	13.6	4.5	25.0	27.9	45.0	991.5	33.1	16.5 D	40.3	2,188	38.3	4,8721	74.6	37,559		
2.9	9.6	4.6	26.9	28.4	56.3	1,034.6	35.6	7.5 D	50.6	1,441	33.1	4,8715	72.9			
2.3	5.3	4.6	31.1	28.7	63.3	970.2	26.6	8 D	52.6	3,456	68.0	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	3,699	69.3	4,8717	72.6			
2.5	6.0	4.6	35.2	28.3	60.8	962.7	27.9	11 D	48.7	4,735	73.1	4,8720	73.2	August	44,300	
2.5	6.3	4.8	40.5	28.0	54.3	910.6	20.8	17.5 D	41.8	2,953	63.8	4,8703	69.6			
2.6	7.4	4.9	43.7	27.8	49.3	948.0	25.9	19 D	40.1	1,395	49.4	4,8693	68.0			
3.7	13.6	5.3	49.5	27.7	47.7	931.1	23.9	34.5 D	22.7	8249	45.5	4,8699	61.3			
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	4,8651	56.9	Sept.	117,304	
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	28.5	1,033.6	38.6	25 D	34.7	2,689	30.2	4,8564	35.2			
4.3	30.6	5.3	64.7	27.1	29.8	1,058.7	44.3	26 D	33.5	3,434	34.8	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	26.1	4,8557	31.9	Oct.	152,716	
4.5	27.9	* 6.2	* 61.7	27.3	33.0	1,135.2	59.0	39 D	27.2	3,883	29.0	4,8538	27.3			
4.0	24.4	* 5.1	* 61.5	27.3	33.0	1,094.1	46.4	29.5 D	29.0	3,543	30.0	4,8540	29.7			
3.6	19.4	* 4.9	* 53.2	27.5	33.0	1,132.3	49.6	27.5 D	30.8	3,014	29.5	4,8549	32.9	Nov.	96,743	
6.5	29.3	* 4.9	* 51.4	27.6	34.1	1,144.0	50.1	31 D	34.2	2,700	34.7	4,8576	41.5			
7.1	32.9	* 4.9	* 48.9	27.2	36.4	1,140.7	54.2	29 D	27.6	2,666	37.1	4,8567	39.7			
5.4	30.3	* 4.9	* 51.3	27.1	27.5	1,077.6	45.3	20 D	33.4	1,630	43.6	4,8554	38.8			
4.8	26.1	* 5.0	* 43.5	27.4	22.7	1,283.9	65.7	4.5 D	71.2	563	48.6	4,8594	44.1	Dec.	34,437	
4.2	26.1	* 4.7	* 46.0	27.8	29.4	1,177.0	55.6	13 P	33.2	213	49.0	4,8612	49.3			
4.0	26.8	4.8	48.6	27.6	36.1	1,107.7	48.1	2.5 D	47.3	836	44.2	4,8592	45.6			
4.9	30.3	* 4.7	* 47.8	27.2	32.3	1,191.3	65.2	11.5 P	64.7		615	4,8604	47.0			
5.5	39.2	* 4.8	* 51.6	27.4	24.9	1,222.4	63.5	5 P	65.1	60	61.7	4,8611	49.0			
6.6	46.1	* 4.8	* 49.3	27.5	32.6	1,202.1	60.3	3.5 P	65.1	2,188	61.7	4,8592	45.0			
7.4	49.3	* 4.9	* 52.2	27.7	35.3	1,015.3	35.9	3.5 P	65.1							

AER 1911 thanks to Stefano Dellavigna



Modern preservation

Log In/Create Account

OPENICPSR Find Data Share Data Repositories

Find Data / Data and Code for: "Indirect Savings from Public Procurement Centralization" / Indirect-Effects-Centralization-main

Data and Code for: "Indirect Savings from Public Procurement Centralization"

Principal Investigator(s): Clarissa Lotti, Lear; Arieda Muço, Central European University; Giancarlo Spagnolo, Site - Stockholm School of Economics; Tommaso Valletti, Imperial College London

Version: V1



[DOWNLOAD THIS FOLDER](#)

Name	File Type	Size	Last Modified
code			06/18/2024 01:15:PM
data			06/18/2024 01:16:PM
output			06/18/2024 01:14:PM
CITATION.CFF	text/plain	862 bytes	06/18/2024 09:14:AM
LICENSE.txt	text/plain	1.2 KB	06/18/2024 09:14:AM
README.md	text/x-web-markdown	6 KB	06/18/2024 09:14:AM
main.sh	application/x-sh	2.4 KB	06/18/2024 09:14:AM

Usage Metrics

Overall Project Metrics

14 Views	3 Downloads	3 Publications
--------------------	-----------------------	--------------------------

Folder/File-Level Metrics

0 Views	0 Downloads
-------------------	-----------------------

[Download Detailed Metrics](#)



Exceptions to the Policy

None



...

... there is a grey zone:

- When data do not belong to researcher, no control over preservation, access!
- Sometimes, **ToU** prevent researcher from revealing metadata (name of company, location)

Transparency again

- However:
 - No exception for need to **describe** access (own and other)
 - No exception for need to fully **describe** processing (possibly with redacted code)

Reproducibility in Economics and beyond





Social Science Data Editors

Improving reproducibility in the social and economic sciences

Data and Code Availability Standard

DCAS The [Data and Code Availability Standard \(DCAS\)](#) is a standard for sharing research code and data, endorsed by [leading journals](#) in social sciences. See <https://datacodestandard.org/> for more information.

DOI [10.5281/zenodo.7436134](https://doi.org/10.5281/zenodo.7436134)

DCAS

Data and Code Availability Standard

[About](#) [Journals](#)

	Data	
1	Data Availability Statement	A Data Availability Statement is provided with detailed enough information such that an independent researcher can replicate the steps needed to access the original data, including any limitations and the expected monetary and time cost of data access.
2	Raw data	Raw data used in the research (primary data collected by the author and secondary data not otherwise available) is made publicly accessible. Exceptions are explained under Rule 1.
3	Analysis data	Analysis data is provided as part of the replication package unless they



Data Editors

- American Economic Association (8)
- Econometric Society (3)
- Canadian Journal of Economics (1)
- Royal Economic Society (2)
- Western Economic Association International (1)
- European Economic Association (1)
- Review of Economic Studies (1)

DCAS

Data and Code Availability Standard

Journals

The following journals endorse the Data and Code Availability Standard.

1. American Economic Journal: Applied Economics  
2. American Economic Journal: Economic Policy  
3. American Economic Journal: Macroeconomics  
4. American Economic Journal: Microeconomics  
5. American Economic Review  
6. American Economic Review: Insights  
7. Canadian Journal of Economics 
8. Econometrica 
9. Econometrics Journal
10. Economic Inquiry 
11. Economic Journal 
12. Journal of Economic Literature  
13. Journal of Economic Perspectives  
14. Journal of the European Economic Association 
15. Quantitative Economics 
16. Review of Economic Studies 
17. Theoretical Economics 



Common policies

<https://social-science-data-editors.github.io/>



Data and Code Availability Standard About Journals

	Data	
1	Data Availability Statement	A Data Availability Statement is provided with detailed enough information such that an independent researcher can replicate the steps needed to access the original data, including any limitations and the expected monetary and time cost of data access.
2	Raw data	Raw data used in the research (primary data collected by the author and secondary data not otherwise available) is made publicly accessible. Exceptions are explained under Rule 1.
3	Analysis data	Analysis data is provided as part of the replication package unless they can be fully reproduced from accessible data within a reasonable time frame. Exceptions are explained under Rule 1.
4	Format	The data files are provided in any format compatible with commonly used statistical package or software. Some journals require data files in open, non-proprietary formats.
5	Metadata	Description of variables and their allowed values are publicly accessible.
6	Citation	All data used in the paper are cited.



A template README for social science replication packages.

The template README provided on this website is in a form that follows best practices as defined by a number of data editors at social science journals.

Template README and Guidance

INSTRUCTIONS: This README suggests structure and content that have been approved by various journals, see Endorsers. It is available as Markdown/txt, Word, LaTeX, and PDF. In practice, there are many variations and complications, and authors should feel free to adapt to their needs. All instructions can (should) be removed from the final README (in Markdown, remove lines starting with > INSTRUCTIONS). Please ensure that a PDF is submitted in addition to the chosen native format.

Overview

INSTRUCTIONS: The typical README in social science journals serves the purpose of guiding a reader through the available material and a route to replicating the results in the research paper. Start by providing a brief overview of the available material and a brief guide as to how to proceed from beginning to end.

Example: The code in this replication package constructs the analysis file from the three data sources (Ruggles et al, 2018; Inglehart et al, 2019; BEA, 2016) using Stata and Julia. Two main files run all of the code to generate the data for the 15 figures and 3 tables in the paper. The replicator should expect the code to run for about 14 hours.

Data Availability and Provenance Statements

INSTRUCTIONS: Every README should contain a description of the origin (provenance), location and accessibility (data availability) of the data used in the article. These descriptions are generally referred to as "Data Availability Statements" (DAS). However, in some cases, there is no external data used



Elsewhere: Political Science

The screenshot shows the Cambridge University Press website for the American Political Science Review. The header includes the Cambridge Core logo and navigation links. The main content area features a search bar with the text 'Search American Political Science Review & Past titles' and a 'Submit your article' button. A sidebar on the left contains a menu with 'Research transparency' selected, which is expanded to show a paragraph: 'This journal believes in the importance of transparent and reproducible research. We require reporting their methodology, for example describing details of study design, sources used and procedures used.'

APSR

The screenshot shows the American Journal of Political Science website. The header includes the AJPS logo and navigation links. The main content area features a section titled 'AJPS Verification Policy' with a paragraph: 'The corresponding author of a manuscript that is accepted for publication in the American Journal of Political Science must provide materials that are sufficient to enable interested researchers to verify all of the analytic results that are reported in the text and supporting materials. The document titled "American Journal of Political Science Guidelines for Preparing Verification Files" provides useful information about what information is needed and how it should be organized. All verification files must be stored in a Dataset within the AJPS Dataverse, on the Harvard Dataverse Network. Note that authors also can make their verification files available elsewhere (e.g., their personal website, other data repositories, etc.) as long as all of the necessary files are included in the Dataset on the AJPS Dataverse.'

Other visible elements include a 'TRANSLATE' section with a language selection dropdown, a search bar, and a 'FOLLOW AJPS VIA EMAIL' section with an email address input field and a 'Follow' button.

AJPS



Elsewhere: Sociology



sociological science

[Articles](#) [For Authors](#)

[Home](#) > [Reproducibility Policy](#)

Reproducibility Policy

Over the last decade, we have witnessed a crisis in science in which many admired research studies were found to be non-replicable. Researchers increasingly recognize that publication itself does not imply that findings are reliable. This has led to widespread questioning of the credibility of social science research. In order to advance the credibility of sociological research, Sociological Science has adopted a reproducibility policy.

Starting with submissions received after April 1, 2023, authors of articles relying on statistical or computational methods are required to deposit replication packages as a condition of publication in *Sociological Science*. Replication packages should include the statistical code and — when legally and ethically possible — the data required to fully reproduce the analysis. Sociological Science hopes other high-impact journals in Sociology will follow suit in setting standards for published work.

In addition to depositing replication packages, papers relying on experimental methods must adhere to the journal's registration requirements outlined in the journal's [Policy on Findings from Experimental Data](#) below.

Under many legitimate circumstances, data cannot legally or ethically be made available to readers. When data are not available, they must explain why in the main text of the paper. In such cases, making code and analysis scripts available is required, unless doing so would violate legal or ethical constraints.

Researchers using qualitative data, such as interviews or participant observation data, are not required to deposit their data. We encourage authors to make qualitative data available when possible, and urge them to consider how such as interview protocols or coding schemes can be shared.

Sociological Science



But!



Elsewhere: Sociology



sociological science

Articles For Authors

Home > Reproducibility Policy

Reproducibility Policy

Over the last decade, we have witnessed a crisis in science in which many admired research studies have been found to be non-replicable. Researchers increasingly recognize that publication itself does not imply that findings are credible. In order to advance the credibility of sociological research, Sociological Science has adopted a reproducibility policy.

Starting with submissions received after April 1, 2023, authors of articles relying on statistical or computational methods are required to deposit replication packages as a condition of publication in *Sociological Science*. Replication packages should include the statistical code and — when legally and ethically possible — the data required to fully reproduce the analysis. Sociological Science hopes other high-impact journals in Sociology will follow suit in setting such standards for published work.

In addition to depositing replication packages, papers relying on experimental methods must adhere to the journal's registration requirements outlined in the journal's [Policy on Findings from Experimental Data](#) below.

Under many legitimate circumstances, data cannot legally or ethically be made available to readers. When data are not available, they must explain why in the main text of the paper. In such cases, making code and analysis scripts available is required, unless doing so would violate legal or ethical constraints.

Researchers using qualitative data, such as interviews or participant observation data, are not required to deposit replication packages. We encourage authors to make qualitative data available when possible, and urge them to describe how such as interview protocols or coding schemes can be shared.



HDSR

Search Dashboard Login or Sign Up

HOME ISSUES SECTIONS COLUMNS COLLECTIONS PODCAST SUBMIT ABOUT MASTHEAD

Issue 5.4, Fall 2023 ... 1 more Published on Oct 27, 2023 DOI 10.1162/99608f92.151c41e3 SHOW DETAILS

Crisis? What Crisis? Sociology's Slow Progress Toward Scientific Transparency

by Kim A. Weeden

Published on Oct 27, 2023

last released 10 months ago

Column Editor's Note: *The quest for transparency, the support of scientific inquiry, is key for credible research, regardless of the discipline. We started the [Reinforcing Reproducibility and Replicability column](#) in [Issue 5.3](#) with a heavy focus on*

3



Benefits

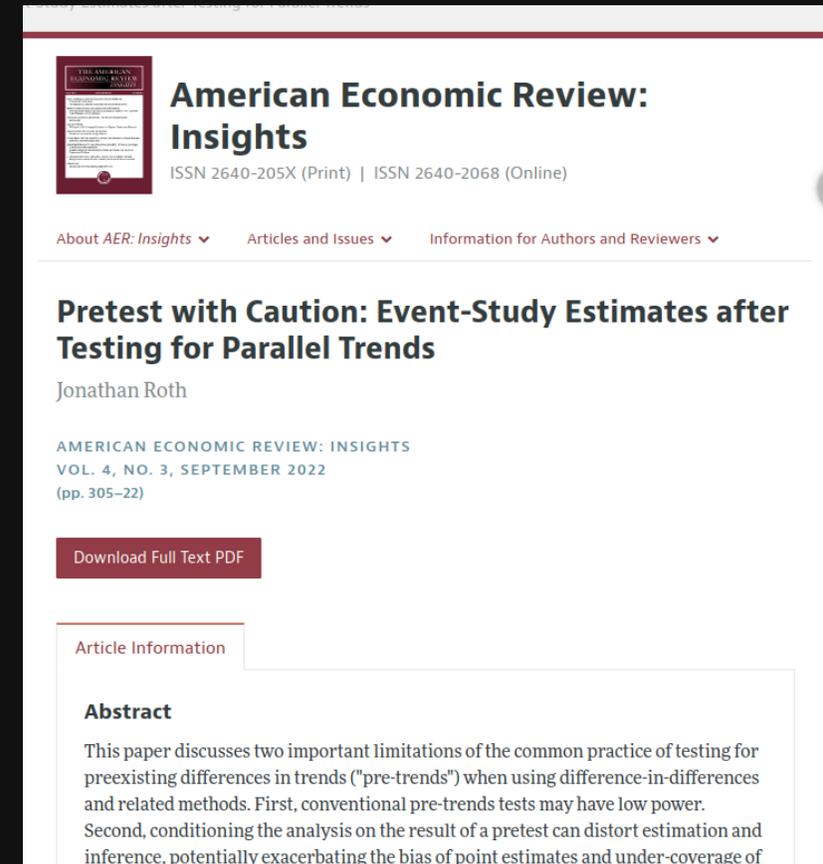


Building on the work of others

Roth, Jonathan. 2022. "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends." American Economic Review: Insights 4 (3): 305–22. DOI:

[10.1257/aeri.20210236](https://doi.org/10.1257/aeri.20210236)

Notes: ***"I exclude 43 papers for which data to replicate the main event-study plot were unavailable."***



The screenshot shows the article page for "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends" by Jonathan Roth in the American Economic Review: Insights. The page includes the journal logo, ISSN information, navigation links, the article title, author name, volume and issue information, a "Download Full Text PDF" button, and an abstract section.

American Economic Review: Insights
ISSN 2640-205X (Print) | ISSN 2640-2068 (Online)

About AER: Insights ▾ Articles and Issues ▾ Information for Authors and Reviewers ▾

Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends
Jonathan Roth

AMERICAN ECONOMIC REVIEW: INSIGHTS
VOL. 4, NO. 3, SEPTEMBER 2022
(pp. 305–22)

Download Full Text PDF

Article Information

Abstract
This paper discusses two important limitations of the common practice of testing for preexisting differences in trends ("pre-trends") when using difference-in-differences and related methods. First, conventional pre-trends tests may have low power. Second, conditioning the analysis on the result of a pretest can distort estimation and inference, potentially exacerbating the bias of point estimates and under-coverage of

Roth 2022

Building on the work of others: dCdH 2020

de Chaisemartin, Clément, and
Xavier D'Haultfœuille. 2020.

“Two-Way Fixed Effects
Estimators with Heterogeneous
Treatment Effects.” *American
Economic Review* 110 (9): 2964–
96. DOI: [10.1257/aer.20181169](https://doi.org/10.1257/aer.20181169)

The results from various other
papers are recomputed to
empirically demonstrate the



The screenshot shows the top portion of a journal article page. At the top left is the journal's logo, 'The American Economic Review'. To its right is the journal title 'American Economic Review' and its ISSN numbers: 'ISSN 0002-8282 (Print) | ISSN 1944-7981 (Online)'. Below this is a navigation bar with three items: 'About the AER', 'Articles and Issues', and 'Information for Authors and Reviewers', each with a dropdown arrow. The main title of the article is 'Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects', followed by the authors' names: 'Clément de Chaisemartin' and 'Xavier D'Haultfœuille'. Below the authors' names is the journal information: 'AMERICAN ECONOMIC REVIEW VOL. 110, NO. 9, SEPTEMBER 2020 (pp. 2964–96)'. A red button labeled 'Download Full Text PDF' is positioned below the journal information. At the bottom of the screenshot, there is a section titled 'Article Information' with a sub-section for the 'Abstract'. The abstract text reads: 'Linear regressions with period and group fixed effects are widely used to estimate treatment effects. We show that they estimate weighted sums of the average treatment effects (ATE) in each group and period, with weights that may be negative. Due to the negative weights, the linear regression coefficient may for instance be negative while all the ATEs are positive. We propose another estimator that solves this issue. In the two applications we revisit, it is significantly different from the linear regression estimator.'



**Transparency
elsewhere**



Transparency outsourced

- Talk to Limor!
- Cornell's R-squared
- cascad
- World Bank

Transparency outsourced

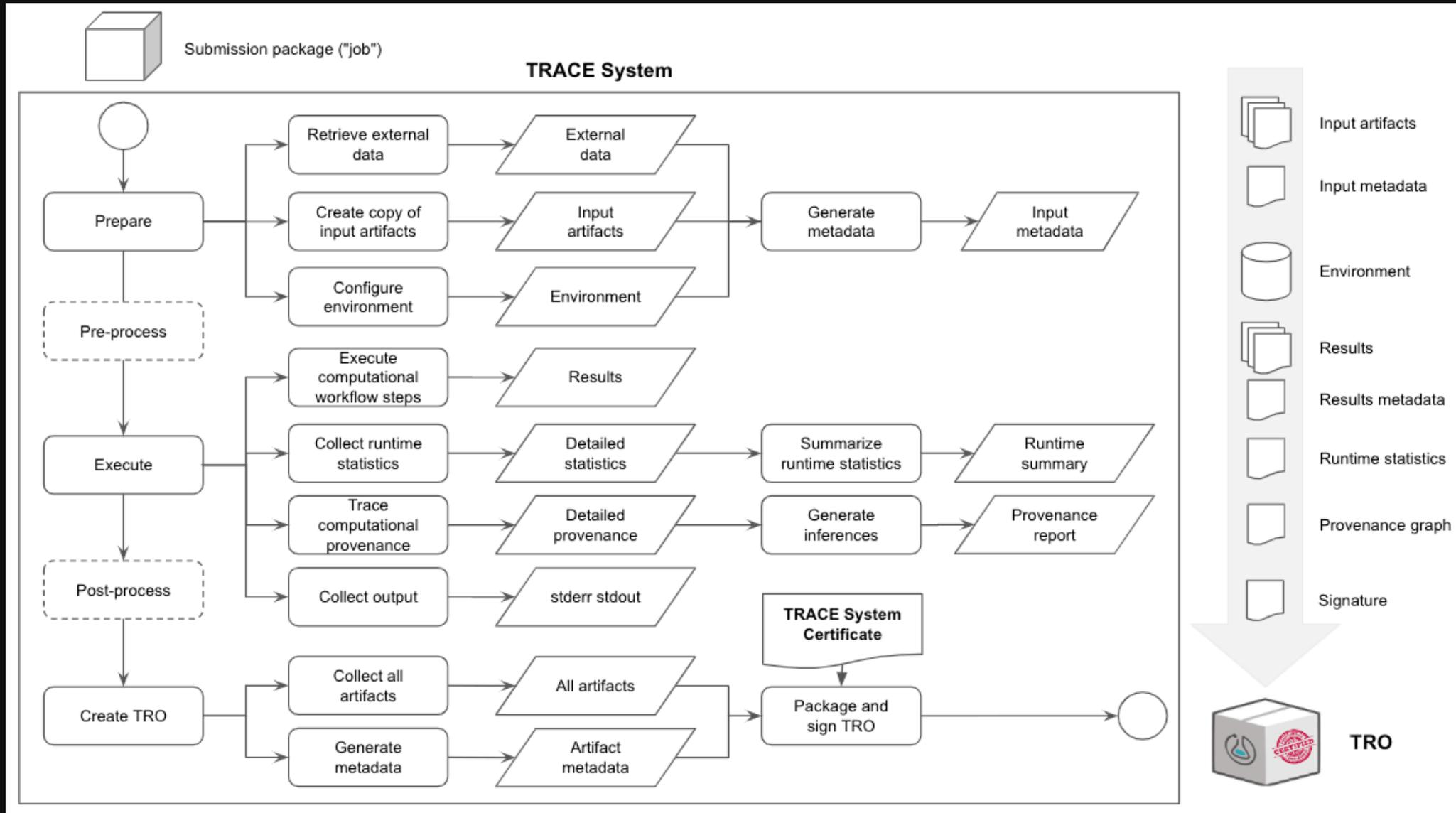
- A third party conducts the reproducibility, not you, not me.
- Need to common understanding, protocols, etc.
- **AEA's protocol**
- We do this about a dozen times per year

Transparency outsourced

Why should I believe the third party?

- Trust
- Transparency
- Common methods

Transparency certified



Transparency certified

- Providing information about the computing platforms themselves, including specific details about how computational transparency is supported.
- Packaging and signing resulting artifacts along with records of their execution using a standard format.



Applications

- Limor, R-squared, cascad, World Bank!
- FSRDC? IRS?
- Meta data?

Footnotes

1.

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles.

Martone M. (ed.) San Diego CA: FORCE11; 2014

<https://www.force11.org/group/joint-declaration-data-citation-principles-final>

2.

National Academies of Sciences, Engineering, and Medicine. 2019.

Reproducibility and Replicability in Science. Washington, DC: The National

Academies Press. <https://doi.org/10.17226/25303>.

3.

Weeden, K. A. (2023). Crisis? What Crisis? Sociology's Slow Progress Toward Scientific Transparency . Harvard Data Science Review, 5(4).

<https://doi.org/10.1162/99608f92.151c41e3>

