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

Lars Vilhuber Michael Stepner

2025-05-26

Follow along



larsvilhuber.github.io/summer-school-qicss-2025/presentation/presentation.html (PDF)



Who are we?



Your instructors







Michael Stepner



Michael Stepner

Assistant professor at the University of Toronto and a Research Principal at Opportunity Insights, a lab based at Harvard University.



Opportunity Atlas

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

Journal of Economic Literature

of literature.

American Economic Journal: Applied

with a focus on empirical

microeconomic issues.



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.

The Journal of Economic Literature

designed to help economists keep

(IEL), first published in 1969, is

American Economic Review: Insights



AER: Insights is designed to be a toptier, 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 Perspectives



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

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

Fconomics

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

Microeconomics American Economic Journal:

American Economic Journal:



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



Iournals

Data Editor of the AEA

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



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

- 13:00 Welcome
- 13:05 Walkthrough
- 13:15 Goals
- 13:30 Technical setup, possible team formation
- 13:45 🔒 Hands-on Exercise: A very imperfect example
- 14:00 Motivation: A tale of ineffective technical collaboration
- 14:05 Day 1: Setting yourself up for reproducibility
- 16:00 End of day one



Tomorrow

- 9:00 Discussion of the "Very imperfect example"
- 9:30 Documenting it all: How to correctly document a replication package (and why!)
- 10:30 Break
- 10:45 How to run Stata! or R! (reproducibly)
- 11:00 Extra: How to install Stata packages
- 11:15 **Topic A** (see Survey)

• 12:00 Break



- 13:00 🔒 When data cleaning is 🔒 critical
- 13:30 **Topic B** (see Survey)
- 14:15 Break
- 14:30 Hands-on: Improving the replication package (very imperfect -> a lot better)
- 15:00 *Hands-on:* Testing it all
- 15:15 Wrap up
- 16:00 Fin.

Best practices?



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

CALL INTEREST RATES ON STOCK EXCHANOR ³ INTEREST RATES ON 60-90 DAY, 2 NAME COMMERCIA PAPER ⁵ BANKS ⁵		CIRCULATION OF DE- FORT CURRENCY		Exchange Chicago	Exchange Rates in N											
Arran	SEASONAL	Avreage	SEASON AL INDEX	Average	SEASONAL INDEX	Average CLEARINGS	SEASONAL INDEX	Average RATE	1899-1908	OUT OF B	AND INTO N ANKS, ^b 1899-1	Y. CITT	STERLING DEMAND	Exchange, Drayes	EXPORTATIO	ON AND INC. 5., 1890-1908
RATE	NUMBER	RATE	NUMBER	PERCENTAGE	44.3	\$1 007 r	NUMBER	(Premium or Discount)	INDEX NUMBER	OUT OF	AMOUNT INTO	SEASONAL	AVERAGE	SEASONAL INDEX	TOTAL EXCESS	TOTAL BYCKE
6.4 3.6	43.4 23.8	5.0 4.7	53.1 41.5	28.6 29.1	64.9 78.8	* 1,253.6 * 1,924.7	60.8 59.6	2.5 P 5 P	64.7			NUMARA	RATE	NUMBER	EXPORTS 000	IMPORTS 000
2.8 2.5	14.9 11.9	4.5	81.9 22.7	30.3	86.9 77.8	* 1,140.0 * 1,190 *	44.0	5 P 10 P	67.4		86,684 6,621	87.9	\$4.8606 4.8657	49.7	Jan. \$32,747	
2.5	11.1 10.1	4.3	22.9	29.9	65.4	1,084.1	\$38.4	2 P	72.1 63.0		7,773	90.7	4.8679	59.4		
2.5	9.8	4.3	92.9 96.5	28.8 28.5	53.6	\$944.0	\$ 32.1	9 D	54.8		4,749	87.6	4,8695	64.1	Feb.	
3.0	13.4	4.6	32.6	28.3	45.5	1,165.7	\$2.6	20 D	38.8		1,436	63.7	4.8696	64.8	13,408	
3.6	19.7	* 4.6	* 34.3	27.9 97.7	43.1 37.0	1,067.9	* 38.2	29.5 D	28.1 95.0		1,157	58.3	4.8697	65.4		Man
3.9	19.2	4.8	39.6	27.9	39.9	1,042.3	33.1	13 D	45.9		604	38.5 30.5	4.8676	63.7		\$ 43,23
3.6	22.0	4.8	38.1	28.0	40.5	1,051.4	35.5	5 D	43.5		716	49.8	4,8665	59.1		0.000
4.0	23.8	4.6	30.7	27.5	39.9	1,119.0	48.0	14 D	44.5		999	53.5	4.8704	65.9	April	
3.0	17.5	4.5	31.9	28.4	50.9	1,123.5	46.7	4 P	52.2		1,903	53.9	4.8711	67.4	10,000	
2.9	15.4	4.4	27.5	28.6 98.3	48.3	1,107.6	43.3	9 D	48.4		2,085	62.1	4.8734	73.6		19
3.5	19.5	4.4	24.5	28.4	48.0	1,175.4	52.7	3.5 D 2.5 P	55.9		594	61,6 56,5	4.8748	78.1	May	10
2.6	13.9	4.3	22.7	28.6	51.6 60.3	1,123.4	48.0	16 P	76.7		\$,952 4 906	65.0	4.8734	74.9	148,048	
2.3	9.6	4.1	17.1	28.8	57.2	908.1	21.4	16 P 10 P	77.3		4,229	74.5	4.8752	75.5	Turn	
2.3	8.0	4.1	15.8	28.7	56.1	1,039.4	37.9	5 P	64.6		3,863	60.9	4.8760	80.9	133,531	
2.5	8.0	4.3	15.3	28.7	57.5	938.7	25.8	4 P 10.5 P	63.6		3,354	66.7	4.8756	81.1 81.0	100000000	
3.6	16.4	4.5	22.0	28.4	53.5	1,013.9	35.4	11.5 P	73.6		3,497	68.5	4.8742	79.0	July	
3.4	13.6	4.5	25.0 96.0	27.9 98.4	45.0	1.034.6	33.1 \$5.6	16.5 D	40.3		1,441	33.1	4.8715	79.0	\$7,359	
2.3	5.3	4.6	31.1	28.7	63.3	970.2	26.6	8 D	39.6	5	3,456	68.0	4.8717	79.6		
2.4	5.6	4.6	33.5	28,7	65.4	924.6	21.1	10.5 D	50.0	1 20	4,735	73.1	4.8720	73.9		Augu
2.5	6.3	4.8	40,5	28.0	54.3	910.6	20.8	17.5 D	41.8		1,395	63.4 57.3	4.8702	69.6		44,3
2.6	7.4	4.9	43.7	27.8	49.3	948.0	25.9	19 D 84.5 D	40.1	8340	9,517	49.4	4.8669	61.3		1 - 2.
3.7	13.6	5.3	49.5	27.7	47.7	956.8	29.0	37.5 D	18.8	1,477		\$3.5	4.8651	56.9		Sep
4.1	20.7	5.3	55.4	27.2	32.8	880.7	19.9	36.5 D	19.1	2,620		99.9 90.9	4.8601	43.7		111,5
4.2	23.4	5.1	57.5	27.0	28.8	1,058.7	44.3	26 D	\$3.5	3,434		94.8	4.8564	35.9 99.0		1
4.9	. 29.6	3.3	63.2	27.5	31.9	1,066.1	36.9	33 D	26.1 97.9	3,489		97.0 39.0	4.8557	31.9		Oct
4.5	27.9	* 6.2	* 61.7	27.3	33.0	1,135.9	46.4	29.5 D	29.0	2,543		52.8	4.8540	97.3 99.7		159,1
4.0	19.4	44.9	4 53.2	27.3 97.5	33.0	1,139.3	49.6	27.5 D	30.8	3,014	1	30.3 29.6	4.8549	32.9		
6.5	29.3	4 4.9	4 51.4	27.6	34.1 S6 4	1,144.0	50.1	31 D 29 D	97.6	2,700	10.1	34.7	4.8567	41.5	1	96,
7.1	32.9	• 4.9	48.9	27.2	27.5	1.077.6	45.3	20 D	36.9	2,666		43.6	4.8554	38.8		
5.4	26.1	• 5.0	• 58.5	27.4	22.7	1,283.9	65.7	4.5 D	71.9	563		48.6	4.8694	44.1		De
4.2	26.1	* 4.7	• 46.0	27.8	36.1	1,177.0	48.1	2.5 D	53.9	913 896		44.9	4.8615	49.3		\$4,0
4.0	26.8	4.8	45.6	27.6	32.3	1,191.3	65.9	11.5 D	64.7		515	52.4	4.8604	45.6	and the second	
5.5	39.2	• 4.8	* 51.6	27.4	24.9	1,992.4	60.8	3.5 P	65.1		2,188	61.7	4.8611	49.0	1	
6.6	46.1	44.8	49.3	27.5	32.8	1.015.3	35.8	3.5 P	65.1				4.8592	45.0	di Canada	1

AER 1911

Increasing broad consensus in academia

- FAIR principles
- Data Citation Principles
- Computational Reproducibility

- Findable
- Accessible
- Interoperable
- Reusable





Data Citation Principles

- FAIR principles
- Data Citation Principles
- Computational Reproducibility

To make it <mark>findable</mark>, <u>citations</u>,

 Give attribution and credit for data.

FORCE11 The Future of Research Communications and e-Scholarship

COMMUNITY - CODE OF CONDUCT GROUPS RESOURCES - NEWS + BLOGS - EVENTS DETCENCED CENTERIA OF IMPORTANCE.

1. Importance

Data should be considered legitimate, citable products of research. Data c should be accorded the same importance in the scholarly record as citatic research objects, such as publications[1].

2. Credit and Attribution

Data citations should facilitate giving scholarly credit and normative and la attribution to all contributors to the data, recognizing that a single style or of attribution may not be applicable to all data[2].

3. Evidence

In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited[3].

4. Unique Identification

A data citation should include a persistent method for identification that actionable, globally unique, and widely used by a community[4].

5. Access

Data citations should facilitate access to the data themselves and to such metadata, documentation, code, and other materials, as are necessary for humans and machines to make informed use of the referenced data[5].



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...

CAL MARKEN

ALL D. MARY

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





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.

Is stored in...

- AEA Data and Code Repository
- Other trusted repositories

OPENICPSR				Find	Data Share Data	a Repositories
Find Data / Data and Code for: "Indirect Sat	vings from Public Procurement Centralization" / I	ndirect-Effects-Centralizati	on-main			
Data and Code for: "Ind "rincipal Investigator(s): © Clarissa Lotti, Lea mperial College London Iersion: © V1	lirect Savings from Public ar, Arieda Muço, Central European University; Gian	Procurement carlo Spagnolo, Site - Stock	Centralization" iholm School of Economics; Tommaso Valletti,			AMERICAN ECONOMIC ASSOCIATION
Name 🖸	File Type 🖸	Size 🖸	Last Modified 🕒			
🖀 code			06/18/2024 01:15:PM		O DOWNLOAD THIS FOLDE	R
🚔 data			06/18/2024 01:16:PM			
🚔 output			06/18/2024 01:14:PM	Usage Metrics 🔞		
CITATION.CFF	text/plain	862 bytes	06/18/2024 09:14:AM	Overall Project Met	rics	
LICENSE.txt	text/plain	1.2 KB	06/18/2024 09:14:AM	14 Views	3 Downloads	3 Publications
README.md	text/x-web-markdown	6 KB	06/18/2024 09:14:AM	Folder/File-Level M	etrics	
main.sh	application/x-sh	2.4 KB	06/18/2024 09:14:AM	v	0 lews Down	0 nloads



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?

- 🐇 Authors at conditional acceptance
- Authors at submission 6 Junior researchers
- of project

- Experienced • 😥 🔞 researchers
- 👦 Undergraduates



Who?

You.



You



- more efficient development
- more efficient collaboration
- more assurance that "everything just works"



- more efficient development across projects
- more efficient response to editors and referrees

... while you are in a new institution, on a new computer, with three courses to prep, and



You



Publish About Browse

🔓 OPEN ACCESS 尨 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 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 <mark>you</mark> 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

CHARLES AND A

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: Al and Big Data but same 1!
- 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/reproducibilityconfidential/
- Guidance for citations: https://social-science-dataeditors.github.io/guidance/addtl-data-citationguidance.html



Extra info

- This document's source: https://github.com/larsvilhuber/summer-school-qicss-2025
- Licensed under CC BY-NC 4.0



Sources

• Images: NYT, Bluesky 1, 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

CALL INTEREST RATES ON STOCK EXCHANGE ^b INTEREST RATES ON 60-90 DAY, 2 NAME COMMERCIAL PAPER ^b BANKS ^b		CIRCULATION OF DE- POSIT CUBRENCYS		THE NEW YORK MONEY MARKET, 1890-1908 Exchange Rates IN Chicago on New York, 1890-1000			1909									
Average	Seasonal Index	Average	Seasonal Index	AVERAGE	SEASONAL INDEX NUMBER	Average Clearings (000,000)	SEASONAL INDEX	AVERAGE RATE (Premium	SEASONAL	AVERAGE	AND INTO N.	Y. CITT	DEMAND	DRAFTS I	EXPORTATIO: GOLD, U. S	N AND INCOM , 1890-1905 FIGURED)
ВАТЕ 6.4	43.4	5.0	53.1	28.6	44.3	\$1,237.5	* 60.8	Discount)	INDEX NUMBER	OUT OF 000	INTO 000	SEASONAL INDEX	AVERAGE RATE	Seasonal Index Number	TOTAL Excess Exports	TOTAL EXCESS IMPORTS
3.6 2.8 2.5 2.4 2.5 3.7 4.12 4.5 6.6 5.1 5.4 4.9 4.5 6.6 6.5 7.4 4.9 4.5 6.6 6.5 7.4 4.9 4.5 6.6 6.5 7.4 4.9 4.5 6.6 6.5 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 4.9 6.5 6.6 7.4 7.4 7.5 7.4 7.5	$\begin{array}{c} 23.6\\ 14.9\\ 11.9\\ 11.1\\ 10.1\\ 9.8\\ 13.4\\ 19.7\\ 22.4\\ 19.7\\ 22.4\\ 19.7\\ 22.4\\ 19.7\\ 22.4\\ 19.7\\ 19.7\\ 19.7\\ 19.7\\ 19.3\\ $	**************************************	\$1.2 22.9 22.1 22.9 22.1 22.9 26.5 34.3 36.7 33.4 30.0 39.6 38.1 36.7 33.4 97.5 96.9 24.3 22.7 19.9 17.1 15.8 15.3 15.3 15.4 22.0 26.9 31.1 33.5 24.0 26.9 31.1 33.5 24.0 26.9 31.1 33.5 24.0 26.9 31.1 33.5 24.0 26.9 31.1 33.5 24.0 26.9 31.1 33.5 24.0 26.9 35.2 40.5 5 4.4 35.2 4.5 5.5 4.4 5.5 4.5 5.5 4.5 5.5 4.5 5.5 4.5 5.5 4.5 5.5 4.5 5.5 5	29.9 29.9 30.3 29.9 29.2 28.8 28.5 28.5 28.5 28.7 28.6 28.8 28.4 28.6 28.8 28.4 28.6 28.8 28.4 28.6 28.8 28.7 28.7 28.7 28.7 28.7 28.7 28.7	78.6 97.7 97.7 97.7 97.7 97.7 97.7 97.7 98.9 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 94.0 95.4 48.0 95.4 49.0 95.4	* 1,224,7 * 1,140,0 * 1,190,5 * 1,084,1 * 1,004,8 * 944,0 * 1,165,7 * 1,067,9 * 1,165,7 * 1,067,9 * 1,107,6 1,293,4 1,119,0 1,119,7 * 1,042,3 1,051,4 1,119,0 * 1,107,6 1,293,4 908,1 1,039,4 908,1 1,035,4 1,035,9 1,046,1 1,035,9 1,144,0 1,145,9 1,144,0 1,144,0 1,145,9 1,140,7 1,190,5 1,200,1 1,200,	• 54.4 • 54.4 • 44.0 • 52.5 • 38.4 • 32.1 • 22.6 • 51.5 • 38.2 • 42.7 33.1 35.5 42.9 46.7 43.3 67.3 52.7 48.0 42.9 46.7 43.3 67.3 52.7 48.0 42.9 46.7 43.3 67.3 52.7 48.0 42.9 46.7 43.3 67.3 52.7 48.0 42.9 46.7 43.3 67.3 52.7 48.0 48.0 42.9 46.7 48.0 49.6 49.0 55.6 55.6 65.9 55.6 60.8 55.8 55.8 55.8 55.8 55.6 60.8 55.8 55.8 55.8 55.8 55.8 55.6 55.6 55.8 55.6 55.6 55.8 55.6 55.6 55.8 55.6 55.6 55.8 55.6 55.8 55.6 55.8 55.6 55.8 55.8 55.6 55.8 55.8 55.6 55.8 55.8 55.6 55.8	5 P 5 P 10 P 5 P 10 P 9 D 90 D 90 D 929.5 D 93.5 D 93.5 D 93.5 D 94.5 D 95.5 D 16.6 P 10.7 P 9 D 2.5 P 16.6 D 11.3 D 94.5 D 95.5 D 91.5 D 92.5 D 93.5 D 94.5 D 95.5 D 90.5 D 91.5 D 92.5 D 93.5 P 94.5 P 95.5 P 95.5 P 95.5 P 95.5 P 95.5 P<	61.4 67.7 67.7 77.1 63.0 54.5 50.7 53.8 58.1 35.9 443.9 443.9 443.9 443.9 663.6 663.7 77.3 77.3 77.3 73.6 50.6 52.6 50.6 52.6 50.7 73.8 73.6 50.0 52.6 53.6 50.6 52.6 53.7 53.8 50.6 52.9 50.6 52.6 53.7 53.8 54.9 55.1 65.1 65.1 65.1	8249 1,417 9,539 3,444 3,683 3,645 3,645 3,645 3,645 2,700 2,665 8,700 8,633 2,700 8,633 2,700 8,633 2,700 8,633 2,133 8,364 8,365 8,364 8,365 8,364 8,365 8,364 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,365 8,366 8,365 8,3666 8,366 8,366 8,366 8,366 8,366 8,366 8,366 8,366 8,366 8,366	80,694 6,691 7,773 6,595 1,436 1,436 1,535 999 809 1,003 1,379 9,952 4,306 4,229 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599 3,597 3,599	87.9 84.9 90.7 87.6 87.6 87.6 87.6 80.3 80.5 80.5 80.5 80.5 80.5 80.5 80.5 80.5	\$4.8006 4.8657 4.8679 4.8679 4.8679 4.8679 4.8692 4.8692 4.8676 4.8692 4.8676 4.8692 4.8676 4.8692 4.8676 4.8692 4.8774 4.8734 4.8754 4.8754 4.8711 4.8712 4.8713 4.8702 4.8669 4.8669 4.8657 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8557 4.8554 4.8556 4.8557 4.8557 4.8556 4.8557 4.8556 4.8557 4.8557 4.8557 4.8556 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.8557 4.85577 4.85577 4.85577 4.85577 4.85577 4.85577 4.855777 4.855777777777	$\begin{array}{c} 42.7\\ 54.7\\ 56.4\\ 64.1\\ 64.1\\ 64.1\\ 64.1\\ 64.2\\ 65.4\\ 65.2\\ 65.2\\ 65.2\\ 65.2\\ 65.2\\ 65.2\\ 65.2\\ 65.2\\ 73.2\\ 75.5\\ 75.5\\ 74.2\\ 75.5\\ 75.5\\ 75.2\\ 74.2\\ 75.5\\ 75.2\\ 73.2\\ 99.2\\ 99.2\\ 99.2\\ 99.2\\ 66.3\\$	Jan, 852,747 Peb, 13,409 May 148,048 June 133,531 July 37,559	March \$ 43,233 August 44,300 Sept. 117,904 Oct. 152,716 Nor. 95,743 Dec. 34,437

AER 1911 thanks to Stefano Dellavigna

Modern preservation

OPENICPS	R			Find	Data Share Dat	Log In/Create Account a Repositories
Find Data / Data and Code for: "Indirec	ct Savings from Public Procurement Centralization"	Indirect-Effects-Centralization	on-main			
Data and Code for: "Ir Principal Investigator(s): • Clarissa Lotti Imperial College London Version: • V1	ndirect Savings from Public i, Lear; Arieda Muço, Central European University; Giar	Procurement ncarlo Spagnolo, Site - Stock	Centralization" holm School of Economics; Tommaso Valletti,			AMERICAN ECONOMIC ASSOCIATION
Name 🖸	File Type 🖸	Size 🖸	Last Modified 06/18/2024 01:15:PM 06/18/2024 01:16:PM			ĒR
CITATION.CFF	text/plain	862 bytes	06/18/2024 01:14:PM 06/18/2024 09:14:AM	Usage Metrics () Overall Project Me	trics	
LICENSE.txt	text/plain text/x-web-markdown	1.2 КВ 6 КВ	06/18/2024 09:14:AM 06/18/2024 09:14:AM	14 Views	3 Downloads	3 Publications
main.sh	application/x-sh	2.4 КВ	06/18/2024 09:14:AM	\	0 /iews Dow	0 nloads

Exceptions to the Policy None



$\bullet \bullet \bullet$

... 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.





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.
		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)

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/

JCΛS

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 beainning 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 Stratements" (DAS). However, in some crast, these is no asternal data used



Elsewhere: Political Science

CAMBRIDGE UNIVERSITY PRESS	문 Products and Se
Cambridge Core Browse ~ Services	s 🗙 Open research 🛩
Home > Journals > American Political Science Review >	Journal information > Journal policies > Research transparency
NO CH	American Political Science Review
A BLANE	Search American Political Science Review & Past titles
	Search within full text
	Submit your article Announcements Subscribe
Journal home Journal information Special Col	llections FirstView articles Latest issue Most cited All issues Open access articles
About this journal 🗸	
Author instructions	Research transparency
Peer review information	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 a
Journal policies	procedures used.

APSR

Home About Contact Subscribe/Join Current Issue



HOME AJPS ARTICLES REVIEWS MANUSCRIPTS

ABOUT AJPS

AJPS Verification Policy

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.

The corresponding author should prepare and upload verification materials to the AJPSD at verse before submitting the final draft of the accepted manuscript. The "American Journal of Political Science Quick Reference for Uploading Replication Files"⁺⁺ provides information about creating a Dataset on the AJPS Dataverse and depositing materials into it. The "American Journal of Political Science Qualitative Data Verification Checklist" and "American Journal of Political Science Qualitative Data Verification Checklist" are designed to help authors ensure that they have provided all necessary files.

When the final draft of the manuscript is submitted, the materials will be verified to confirm that they do, in fact, reproduce the analytic results reported in the article. Publication in the *American Journal of Political Science* is contingent upon provision of complete verification materials and successful verification of their content.

AJPS articles that do not contain empirical data analyses are not subject to this verification policy. In some limited circumstances, an author may request an exemption

AJPS

TRANSLATE Select Language Powered by Google Translate Search this website... FOLLOW AJPS VIA EMAIL

Enter your email address to receive notifications of new AJPS Author Summaries and posts from the editors. Email Address



Current Issue – October 2024

AJPS Virtual Issue of Recently Top Cited Papers

The American Journal of Political Science (AJPS) is the flagship journal of the Midwest Political Science Association and is published by Wiley.



Elsewhere: Sociology



Articles For Authors

Home > Reproducibility Policy

Reproducibility Policy

Over the last decade, we have witnessed a crisis in science in which many admired research studie non-replicable. Researchers increasingly recognize that publication itself does not imply that finding questioned the credibility of social science research. In order to advance the credibility of sociologica has adopted a reproducibility policy.

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

In addition to depositing replication packages, papers relying on experimental methods must adhere 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. data available, they must explain why in the main text of the paper. In such cases, making code and required, unless doing so would violate legal or ethical constraints.

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

Sociological Science



But!



Elsewhere: Sociology



Articles For Authors

Home > Reproducibility Policy

Reproducibility Policy

Over the last decade, we have witnessed a crisis in science in which many admired research studie non-replicable. Researchers increasingly recognize that publication itself does not imply that finding questioned the credibility of social science research. In order to advance the credibility of sociologica has adopted a reproducibility policy.

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

In addition to depositing replication packages, papers relying on experimental methods must adhere 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. data available, they must explain why in the main text of the paper. In such cases, making code and required, unless doing so would violate legal or ethical constraints.

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



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 <u>Reinforcing</u> <u>Reproducibility and Replicability column</u> in <u>Issue 5.3</u> with a heavy focus on

Sociological Science



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

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



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
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

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



Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects

Clément de Chaisemartin

Xavier D'Haultfœuille

AMERICAN ECONOMIC REVIEW VOL. 110, NO. 9, SEPTEMBER 2020 (pp. 2964–96)

Download Full Text PDF

Article Information

Abstract

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.

d

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

