I wanted to highlight a recently (2020-04-30) published article in the AEJ:Economic Policy as an example of reproducibility checking when data is restricted-access:

> Leung, Pauline, and Christopher O'Leary. 2020. "Unemployment Insurance and Means-Tested Program Interactions: Evidence from Administrative Data." American Economic Journal: Economic Policy, 12 (2): 159-92. https://doi.org/10.1257/pol.20170262

Unemployment Insurance and Means-Tested Program Interactions: Evidence from Administrative Data

Pauline Leung

Christopher O'Leary

Download Full Text PDF

Article Information   Comments (0)

Abstract

We study the ways in which unemployment insurance (UI) benefits interact with other elements of the social safety net around job losses. We exploit a cutoff for UI eligibility, based on a workers' highest quarterly earnings in the past year, to generate quasi-experimental variation in UI receipt. We find that UI receipt cuts welfare (TANF) receipt by half among low-earning UI applicants but has no impact on SNAP or Medicaid usage. However, because welfare participation is low in this population, overall crowdout is small. In the quarter following layoff, UI increases total income by 55 percent (including labor earnings and transfers).

This manuscript was part of our earliest reproducibility checks - it was assigned to us in early April 2019. But with one caveat that I'll hold till the end, it is an excellent example of a well-documented, and as it turns out, reproducible article. Even if the data is confidential and restricted-access!

The data underlying the analysis is administrative data from various state agencies in Michigan, and it would have been difficult and time-consuming for us replicators to request access.

tables in the paper. It includes only datasets that do not contain confidential program participant information, which are restricted use. Researchers interested in obtaining similar Michigan administrative data should submit Freedom of Information Act (FOIA) requests to the Michigan Unemployment Insurance Agency (https://www.michigan.gov/uia/0,4680,7-118-72784-358492--,00.html) and to the Michigan Department of Health and Human Services (https://www.michigan.gov/mdhhs/0,5885,7-339-72666-357860--,00.html).

We reached out to the authors and asked whether they would be able to add a 3rd-party to their research protocol - either a member of our team or somebody at their institution (@Upjohninstitute) who was not involved in the research.

The authors connected us with a research associate at their institution who had legal access to the data. The procedure happened, as much as possible, at arms-length: the 3rd-party replicator followed only written instructions that are publicly available, as per our protocol.

In fact, all information and code came from the researchers' replication archive at openICPSR, which is now available at https://doi.org/10.3886/E109704V1. However, the replicator had access prior to publication of the archive, and what you see now is not what they had access to.

Data and Code for: Unemployment Insurance and Means-Tested Program Interactions: Evidence from Administrative Data

Principal Investigator(s): ❷ Pauline Leung, Cornell University; Christopher O'Leary, W.E. Upjohn Institute

Version: ❷ V1

Name ⊡	File Type ⊡	Size ⊡	Last Modified ⊡
📁 ado			06/15/2019 10:13:AM
📁 aux_data			05/11/2020 12:33:PM
📁 log			06/15/2019 10:14:AM
📁 output			06/15/2019 10:14:AM
📁 programs			06/15/2019 10:14:AM
📁 restricted_data			06/15/2019 10:14:AM
📁 results			06/15/2019 10:14:AM
📄 readme.pdf	application/pdf	40.2 KB	06/15/2019 06:13:AM

◉ DOWNLOAD THIS PROJECT

Usage Metrics ❷

Overall Project Metrics

4	0	1
Views	Downloads	Publications

Download Detailed Metrics

Published Versions

V1 [2020-05-11]

Export Metadata

Dublin Core

Project Citation:

Leung, Pauline, and O'Leary, Christopher. Data and Code for: Unemployment Insurance and Means-Tested Program Interactions: Evidence from Administrative Data. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-05-11. https://doi.org/10.3886/E109704V1

That's because the replicator identified and suggested multiple (small) edits to README and programs, to make the replication easier for subsequent replicators.

They also identified (minor) various numerical discrepancies, typically at the 3rd decimal place, that we tracked down to differences in the versions of the Stata package rdrobust. After discussions with rdrobust authors and manuscript authors, the replication package now contains the exact version of rdrobust used for the analysis.

Overall, the report received from the 3rd party replicator was useful confirmation that the code that anybody can see would work well on the data that few people can see, and materially improved the replication package.

Take-away message: While we cannot do reproducibility checks on restricted-access data in a timely fashion for all papers that use restricted-access data, there are still many opportunities to do so, and we will always attempt to do so.

In fact, since that paper was reproduced, we have been able to make similar attempts at reproduction for about a dozen other papers, for which we reached out to various individuals or institutions. We have also signed NDA or DUA for various datasets, to obtain copies ourselves. Not all attempts succeed, but expect to see more such verifications in the future.

Caveat I alluded to: We were still refining our various checks and guidance documents. While the authors did an excellent job citing various data sources (in the online appendix), we did not ask them to add a data citation for the core confidential data. That would be different today.