# AEJPol-2019-xxxx.R1 "PAPER TITLE" Validation and Replication results

You may want to consult Unofficial Verification Guidance for additional tips and criteria.

[NOTE] This is an amalgam of multiple reports, and updated to reflect template and guidance as of February 2022.

#### **SUMMARY**

Thank you for your replication archive. All tables and figures are replicated, with one minor discrepancy in Table 1.

Conditional on making the requested changes to the manuscript and the openICPSR deposit prior to publication, the replication package is accepted.

In assessing compliance with our Data and Code Availability Policy, we have identified the following issues, which we ask you to address:

#### Action Items (manuscript)

(these are handled during copy-editing, and involve no further interaction with the Data Editor)

- [REQUIRED] Please add data citations to the article. Guidance on how to cite data is provided in the AEA Sample References and in additional guidance.
- [REQUIRED] Please adjust your tables to account for the noted numerical discrepancies, or explain (in the README) discrepancies that a replicator should expect.

#### Action Items (openICPSR)

(these are handled prior to publication, and are verified by the Data Editor, but no code is re-run).

- [REQUIRED] Please provide debugged code, addressing the issues identified in this report.
- [REQUIRED] Please provide missing code (appendix table)
- [REQUIRED] Please adjust your tables to account for the noted numerical discrepancies, or explain (in the README) discrepancies that a replicator should expect.
- [SUGGESTED] The authors should check if the three surveys are cataloged and stored in the World Bank Microdata catalog. If yes, referencing the relevant entries from the Microdata catalog is required.
- [REQUIRED] openICPSR should not have ZIP files visible. ZIP files should be uploaded to openICPSR via "Import from ZIP" instead of "Upload Files". Please delete the ZIP files, and re-upload using the "Import from ZIP" function.
- [REQUIRED] Please amend README to contain complete requirements.

## Data description

#### **Current Employment Statistics**

- Dataset is not provided, but a link is provided in the README
- Access conditions are not described. It turns out, the data are in the public domain. This might warrant being identified in the report.
- The data are cited in the references section of the manuscript and the README. Data citation:

Bureau of Labor Statistics. 2000–2010. "Current Employment Statistics: Colorado, Total Nonfarm, Seasonally adjusted - SMS080000000000001." United States Department of Labor. http://data.bls.gov/cgi-bin/surveymost?sm+08 (accessed February 9, 2011).

Surveys by World Bank: Baseline survey, 3-year (200x) and 6-year (201x) follow-ups (PII removed)

- Data Source: World Bank and the authors' survey
- The World Bank collected both a baseline survey and initial follow-up survey to assess short-run program impacts.
- Note that the raw survey datasets the authors have published have been modified slightly from the original raw data in order to remove PII. In particular, they have removed the recorded date of birth for all survey respondents (year and month of birth are still included).
- The original study describing the first two surveys was cited in the manuscript. It appears that the last follow-up survey is conducted by the authors. Access information is detailed in the readme.

NOTES: The citation in the manuscript is to a description of the first two waves, not to the actual data.

[SUGGESTED] The authors provide data from Waves I and Wave II, and Wave III (conducted by the authors). Presumably, all three surveys are cataloged and stored in the World Bank Microdata catalog. Referencing the relevant entries from the Microdata catalog is strongly suggested, as it provides additional metadata and findability.

[REQUIRED] Please add data citations to the article. Guidance on how to cite data is provided in the AEA Sample References and in additional guidance.

#### **Analysis Data Files**

- No analysis data file mentioned
- Analysis data files mentioned, not provided (explain reasons below)
- Analysis data files mentioned, provided. File names listed below.

./data\_intermediate/anon\_survey4.dta

## Data deposit

#### Requirements

- README is in TXT, MD, PDF format
- openICPSR deposit has no ZIP files
- Ittle conforms to guidance (starts with "Data and Code for:" or "Code for:", is properly capitalized)
- Authors (with affiliations) are listed in the same order as on the paper

[REQUIRED] openICPSR should not have ZIP files visible. ZIP files should be uploaded to openICPSR via "Import from ZIP" instead of "Upload Files". Please delete the ZIP files, and re-upload using the "Import from ZIP" function.

Detailed guidance is at https://aeadataeditor.github.io/aea-de-guidance/.

#### Deposit Metadata

- IEL Classification (required)
- Manuscript Number (required)
- Subject Terms (highly recommended)
- Geographic coverage (highly recommended)
- Time period(s) (highly recommended)
- Collection date(s) (suggested)
- Universe (suggested)
- Data Type(s) (suggested)
- Data Source (suggested)
- Units of Observation (suggested)
- [NOTE] openICPSR metadata is sufficient.

#### However,

- [SUGGESTED] We suggest you update the openICPSR metadata fields marked as (highly recommended), in order to improve findability of your data and code supplement.
- [SUGGESTED] We suggest you update the openICPSR metadata fields marked as (suggested), in order to improve findability of your data and code supplement.

For additional guidance, see https://aeadataeditor.github.io/aea-de-guidance/data-deposit-aea-guidance.html.

## Code description

- Each figure in the article corresponds to specific lines in the provided programs; see "code-check.xlsx" (end of this report) and "Readme.pdf" for details.
- The main analysis programs are written in Stata and the could be executed through the given master program MASTER.do.

• The code segment that changes the directories in every programs does not run properly. Manual changes are needed for individual .do file.

- Carthago-Lion p-values is calculated in the program, which takes a long time (could be more than a day in some computers as mentioned in the Readme).
- No code is provided for Appendix Figure 1 and 3 since they they are not empirical.
- Code for Appendix Table C-2 could not be identified.

[SUGGESTED] Please specify hardware requirements, and duration (execution time) for the last run, to allow replicators to assess the computational requirements.

#### Data checks

- · All datasets are present for this manuscript.
- Data can be read using Stata, and have data variable labels
- Ran PII and results are reported in "pii\_stata\_output.csv".

## **Stated Requirements**

- No requirements specified
- Software Requirements specified as follows:
  - Stata
- Computational Requirements specified as follows:
  - Cluster size, etc.
- Time Requirements specified as follows:
  - Length of necessary computation (hours, weeks, etc.)
- Requirements are complete.

## Missing Requirements

- Software Requirements
  - ∘ ✓ Stata
    - Version
    - Packages go here
- Computational Requirements specified as follows:
  - Cluster size, disk size, memory size, etc.
- **I** Time Requirements
  - Length of necessary computation (hours, weeks, etc.)

[SUGGESTED] While running a static code scanner for Stata, we identified possible Stata packages used in your code. Please verify, and adjust requirements accordingly.

Package	Rank on whatshot
estout	2
reghdfe	5

[REQUIRED] Please amend README to contain complete requirements.

## Computing Environment of the Replicator

- CISER Shared Windows Server 2016, 256GB, Intel Xeon E5-4669 v3 @ 2.10Ghz (3 processors)
- Stata/MP 16

## Replication steps

- 1. Downloaded data and code from the openIPCSR repository.
- 2. Fixed directory errors in the all Stata programs and changed the package fmttable to frmttable in MASTER.do.
- 3. Ran MASTER. do per Readme.
- 4. Reproduced all tables and figures.

[SUGGESTED] Please add a setup program that installs all packages as noted above. Please specify all necessary commands. An example of a setup file can be found at https://github.com/gslab-econ/template/blob/master/config/config\_stata.do

[REQUIRED] Please provide debugged code, addressing the issues identified in this report.

# **Findings**

## Code Check Table

Figure/Table #	Program	Line Number	Replicated?
Table 1	table1.do		minor difference in F-statistics
Table 2	table2.do		Yes
Table 3	table3.do		Yes
Table 4	table4.do		Yes
Table 5	table5.do		Yes
Table 6	table6.do		Yes
Table 7	table7.do		Yes
Appendix Table 1	appendix_table1.do		Yes
(abbreviated)			

Figure/Table #	Program	Line Number	Replicated?
Appendix Table 29	appendix_table29.do		Yes
Appendix Table C-2	not found		No
Appendix Figure 1	not empirical		N/A
Appendix Figure 2	appendix_figure2.do		Yes
Appendix Figure 3	not empirical		N/A
Appendix Figure 4	appendix_figure4.do		Yes
Appendix Figure 5	appendix_figure5.do		Yes

Minor directory issues in the all do-files (the automated directory changing code does not work properly).
Manual changes are needed

#### **Tables**

- Table 1: Mostly reproduced (minor difference in F-statistics)
  - Paper: Table 1, Col. 2: 456.783
  - Reproduced: Table 1, Col 2: 456.812
- Table 2-7: Fully replicated
- Appendix Table 1-29: Fully replicated
- Appendix Table C-2: Not found

[REQUIRED] Please adjust your tables to account for the noted numerical discrepancies, or explain (in the README) discrepancies that a replicator should expect.

### **In-Text Numbers**

- ${f \emph{M}}$  There are no in-text numbers, or all in-text numbers stem from tables and figures.
- ☐ There are in-text numbers, but they are not identified in the code

## Classification

- In full reproduction
- If full reproduction with minor issues
- partial reproduction (see above)
- not able to reproduce most or all of the results (reasons see above)

## Reason for incomplete reproducibility

INSTRUCTIONS: mark the reasons here why full reproduciblity was not achieved, and enter this information in JIRA

- Discrepancy in output (either figures or numbers in tables or text differ)
- Bugs in code that were fixable by the replicator (but should be fixed in the final deposit)
- Code missing, in particular if it prevented the replicator from completing the reproducibility check
  - Data preparation code missing should be checked if the code missing seems to be data preparation code
- Code not functional is more severe than a simple bug: it prevented the replicator from completing the reproducibility check
- Software not available to replicator may happen for a variety of reasons, but in particular (a) when the software is commercial, and the replicator does not have access to a licensed copy, or (b) the software is open-source, but a specific version required to conduct the reproducibility check is not available.
- Insufficient time available to replicator is applicable when (a) running the code would take weeks or more (b) running the code might take less time if sufficient compute resources were to be brought to bear, but no such resources can be accessed in a timely fashion (c) the replication package is very complex, and following all (manual and scripted) steps would take too long.
- Data missing is marked when data *should* be available, but was erroneously not provided, or is not accessible via the procedures described in the replication package
- Data not available is marked when data requires additional access steps, for instance purchase or application procedure.