Replication instructions for The Null Result Penalty"

Felix Chopra, Ingar Haaland, Christopher Roth, Andreas Stegmann

June, 2023

Overview

The code in this replication package constructs the analysis file from the two data sources (Chopra et al, 2023; Andre and Falk, 2021) using Stata and R. The main Stata do-file runs all of the code to generate the 5 figures and 16 tables in the paper. The replicator should expect the code to run for less than an hour.

Data Availability and Provenance Statements

The main data for this paper was collected by the authors as part of a set of two survey experiments and are included in this replication package. The replication package also includes the set of experimental instructions and questionnaires for all experiments. In addition, we manually collected data and draw on data from Andre and Falk (2021) that was generously shared with us. We elaborate on the data availability of the external data below.

Statement about Rights

- ⊠ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- [] I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

- ☐ All data **are** publicly available.
- ⊠ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on each Data Source

Data.Name	Data.Files	Location	Provided	Citation
"Main experiment"	main_study_raw.dta	data/raw/	TRUE	Chopra et al (2023)
"Mechanism experiment"	mechanism_study_raw.dta	data/raw/	TRUE	Chopra et al (2023)
"External CV data"	expert_demographics.dta	data/	TRUE	Chopra et al (2023)
"Population stats"	composition.dta	data/	TRUE	Andre and Falk (2021)
"Researchers by JEL"	jel.dta	data/	TRUE	Andre and Falk (2021)

The data for the main experiment, the mechanism experiment, and the external CV data of our study participants are included in the replication package.

There are three variables in the "External CV data" file that we do not include in this replication package to protect the anonymity of our respondents. These variables are `top5count` (number of top 5 publications), `gshindex` (GoogleScholar h-index), and `gscitation` (GoogleScholar citation count). This information was collected from respondents' GoogleScholar pages after our experiments. However, our analysis only uses two transformed variables: a) whether the respondent has any top 5 publications (`anytop5`) and b) whether the respondent has above-median citations (`highcite`). We include these transformed variables in the version of the dataset that is included in the replication package, but do not include the continuous variables to ensure that respondents cannot be identified. The transformed variables are needed to generate Figure 2 and Figure A1. Moreover, Table 1 and Table A4 report summary statistics (mean and median) for the variables `top5count`, `gshindex`, and `gscitation`. As we cannot share the underlying continuous variables, we encoded the summary statistics in the Stata do-files that generate these tables.

Replicators that wish to conduct robustness checks of our results that involve other transformations of the variables `top5count`, `gshindex`, or `gscitation` may contact us. We will then assist them and provide the transformed variables.

The population summary statistics for the universe of research economists at the top 200 institutions according to RePEc were generously shared by Andre and Falk (2021). The data are available as part of the replication package. They can be used by other researchers provided they include a citation to Andre and Falk (2021).

Code for data cleaning and analysis is provided as part of the replication package.

Computational requirements

Software Requirements

- Stata (code was last run with version 16)
  - `estout`
  - `blindschemes`
  - `rwolf2`
  - `addplot`
  - `cdfplot`
  - `reghdfe`
  - `coefplot`
  - `grstyle`
  - `winsor`
  - `winsor2`
  - The program "`main.do`" will install all dependencies.
- R 4.2.2
  - `tibble`
  - `plyr`
  - `dplyr`
  - `Hmisc`
  - `tidyr`
  - `anesrake`
  - `weights`
  - the file "`weights_R/calculate_weights.R`" will install all dependencies (latest version) before performing the calculations.

Memory and Runtime Requirements

Summary   Approximate time needed to reproduce the analyses on a standard (2023) desktop machine:

- ☐ <10 minutes
- ☒ 10-60 minutes
- ☐ 1-2 hours
- ☐ 2-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Details   The code was last run on a **12-core Intel-based laptop with 32 GB of RAM and with MacOS version 13.4**. The computations took about 20 minutes.

Description of programs/code

All programs below are called from `main.do` in the `code`folder of the replication package: - The file `setup.do` sets parameters and the PATH variables. Replicators have to change the path variables here before calling `main.do`. - Programs in `code/cleaning` will clean the raw data collected as part of our experiments. - Programs in `code/figures` will create the figures for the paper. - Programs in `code/tables` will create the tables for the paper.

Calculations in R are performed by the file `weights_R/calculate_weights.R`.

Instructions to Replicators

- Edit `code/setup.do` to adjust the default path (lines 14 to 29).
- Run `code/main.do` once on a new system to set up the working environment and generate all figures and tables.

Details

- `code/setup.do`: will set up paths, local and global macros, and adjust the settings relevant for the production of graphics.
- `code/main.do`: This program calls `setup.do`, installs all necessary libraries, calls the do-files for data cleaning, and then calls the individual do-files that generate all tables and figures for the paper.

List of tables and programs

The provided code reproduces:

- ☐ All numbers provided in text in the paper
- ☒ All tables and figures in the paper
- ☐ Selected tables and figures in the paper, as explained and justified below.

Figure/Table	Program	Output file
Table 1	code/tables/table_1.do	tables_1.tex
Table 3	code/tables/table_3.do	tables_3.tex
Table 4	code/tables/table_4.do	tables_4.tex
Table 5	code/tables/table_5.do	tables_5.tex
Table A1	code/tables/table_A1.do	tables_A1.tex
Table A2	code/tables/table_A2.do	tables_A2.tex
Table A3	code/tables/table_A3.do	tables_A3.tex
Table A4	code/tables/table_A4.do	tables_A4.tex
Table A5	code/tables/table_A5.do	tables_A5.tex
Table A6	code/tables/table_A6.do	tables_A6.tex
Table A7_A8	code/tables/table_A7_A8.do	tables_A7_A8.tex
Table A9	code/tables/table_A9.do	tables_A9.tex
Table A10	code/tables/table_A10.do	tables_A10.tex

Figure/Table	Program	Output file
Table A11	code/tables/table_A11.do	tables_A11.tex
Table A12	code/tables/table_A12.do	tables_A12.tex
Figure 2	code/figures/figure_2.do	figures_2.tex
Figure 3	code/figures/figure_3.do	figures_3.tex
Figure A2	code/figures/figure_A2.do	figures_A2.tex
Figure A3	code/figures/figure_A3.do	figures_A3.tex

References

Andre, P. and Falk, A. (2021). "What's worth knowing? Economists' opinions about economics", ECONtribute Discussion Paper.

Chopra, F., Haaland, I., Roth, C. and Stegmann, A. (2023). "The Null Result Penalty".