Readme for the replication package of **Do Financial Concerns Make Workers Less Productive?**

Supreet Kaur, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach

February 7, 2023

Abstract

This is the replication package for "Do Financial Concerns Make Workers Less Productive?," by Supreet Kaur, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach. This folder contains all the data and code necessary for replicating the tables and figures in the paper, as well as English translations of the survey instruments used in data collection. The data files are in Stata format and replication code was written in Stata and R. The replication time will depend on whether the randomization inference iterations are run for the adjusted p-values in Table A8: replicating without takes about 5 minutes; replicating with can take much longer.

Contents

1 Top-folder structure **2**

2 Replication instructions **3**

3 Dataset information, availability, and provenance **3**

4 Dataset list **4**

5 Computational requirements **4**

6 Description of programs/code **5**

1 Top-folder structure

- **ado** – Contains ado files as supplements for the MHT programs in Table_A8.do
- *codebook.xlsx* – Contains variable-level metadata, including full value labels.
- *compile_tables.tex* – Run to compile the tex files for all tables except for A8 from the directory.
- **Data** – Contains the replication data. Described in more detail in section 3
- *main.do* – Used to run the replication. Described in more detail in section 2
- **Output** – Contains folders for tables and figures from both the main paper and the appendix.
  - Appendix
    - Figures
    - Tables
  - Main
    - Figures
    - Tables
- **Scripts** – The replication scripts, separated by whether they replicate tables and figures from the main paper or from the appendix.
  - *0_generate_variables.do*
  - Appendix
  - Main
- **Survey instruments** – The instruments used to collect the study's survery data.

2 Replication instructions

1. Open main.do and follow the instructions below (also in main.do)
2. Set the path to the data folder in the lines marked "SET DIRECTORY HERE" (Lines 19 and 46 or 56)
3. Set the globals "MHT", "reps", and "rsour"; the first takes 1 if you would like to replicate the multiple hypothesis testing table (A8) and 0 if not. The second sets the number of replications for the multiple hypothesis test corrections. The final takes 1 if you would like to run the r code that replicates Figure 1 as well, and 0 (or other) if not.
4. Run main.do

Notes:

- The user-written commands that are necessary to replicate are listed in the loop in section 2 of the main do-file. The loop will check whether your Stata instance has the package installed, and install it if not. If you do not want to install any of the packages, comment this section out. However, replicability will be affected.
- The package currently uses a stata package, RSource, in order to run the r scripts necessary to replicate within Stata. If you would not like to run the R scripts in stata, set global rsour to 0.

3 Dataset information, availability, and provenance

This package contains data from an RCT conducted in Odisha, India, between 2017 and 2018.

The data comes from 4 sources:

1. A baseline survey including basic demographics such as age, education, measures of income and wealth, and information about outstanding loans and financial worries.
2. An endline survey conducted at the end of each round, which collected information about financial worries as well as expenditure patterns and food consumption over the last 3 to 4 days
3. A short survey on day 10 or 11 asking workers about what they thought about while working that day.
4. Data on productivity and attentiveness gathered from the output of the workers.

The data from the above sources is publicly available at the QJE dataverse under license CC0. The available data is cleaned analysis data; variables are labelled, and where applicable, have had generations marked (many variables are standardized, etc.).

For more information on the data sources for this replication package, see sections 3 and 4 of the associated paper and section D of its appendix.

4 Dataset list

The following datasets are included base-datasets (i.e., not generated by running the replication code):

- financial_strain (.dta)

The following datasets are included generated-datasets (i.e., replaced each time the full replication code is run):

- analysis_all (.dta)
- analysis_expenditure (.dta)
- analysis_main (.dta)
- TEMP1 (.dta)
- TEMP2 (.dta)

5 Computational requirements

All programs were written in Stata and R. The following non-base packages are required:

Table 1: Software requirements

Stata	**R**
estout	rio
pdslasso	SnowballC
rlasso	wordcloud2
reghdfe	RColorBrewer
ftools	tm
wyoung	hunspell
rsource	tidytext
	ggplot2
	tibble
	tidyverse
	htmlwidgets
	webshot

The code was last run in Stata 14 on an Apple M1 computer with MacOS version 12.2.1. Replicating without the randomization inference iterations took approximately 5 minutes. A run to fully replicate the paper requires a rep set of 2,000 and takes approximately 3.5 hours on the machine described above.

6 Description of programs/code

With the exception of Table_A8.do (which produces a folder of tex files), each do-file in the "scripts" folder reproduces exactly one table or figure from the paper, outputting the results into a similarly named file in the "Output" folders. The scripts with the "A" prefix produce appendix tables/figures, and the results output to the "Output/Appendix" folder; the do-files without the "A" prefix reproduce the tables from the main paper, and output results to the "Output/Main" folder.

main.do runs set-up, sets globals, and calls all scripts.