README for *”Selfish Corporations”*

This replication package contains all data and code required to replicate the tables and figures of the paper:

Colonnelli, E., Gormsen, N., & McQuade, T. (2022). Selfish Corporations. Review of Economic Studies.

Two code files in the main folder run all Stata (Master\_Stata.do) and R (Master\_R.R) codes in the required order to generate all figures and tables in the main paper. These files use the scripts in the folder *Code* and the data in the folder *Data*. The files Master\_Stata.do and Master\_R.R create all output (tables and figures) and place it in the *Output* folder.

1. Data

The main data sources used in the paper were the four surveys we ran; one in May 2020, one in October 2020, and one in June 2022. Additionally, we also use the 2019 U.S Current Population Survey (CPS) ([Flood et al. (2022)](#_Bibliography)) for some analyses.

* 1. **Data description/folders.** All the data used in the paper is provided by the authors in the replication package. There are three main sub-folders within the *Data* folder:
     1. *Data/Raw.* Contains the raw data collected from the surveys, each file contains the data described below:
        + **Data/Raw/Final\_Data\_May2020\_Qualtrics.dta:** contains all the data from the main survey conducted in May 2020.
        + **Data/Raw/Final\_Data\_October2020\_Qualtrics.dta:** contains all the data from the survey conducted in October 2020.
        + **Data/Raw/June2022\_WCTM\_Option\_1\_Qualtrics.csv & Data/Raw/** **June2022\_WCTM\_Option\_2\_Qualtrics.csv :** contains all the data from the survey conducted in June 2022.
     2. *Data/Cleaned.* Contains the clean data, used in the analyses generated by the code mentioned in Section [2.1](#_bookmark0). The full cleaned data is described below:
        + **Data/Cleaned/Final\_Data\_May2020.dta:** clean data from the main survey conducted in May 2020.
        + **Data/Cleaned/Final\_Data\_October2020.dta:** clean data from the survey conducted in October 2020.
        + **Data/Cleaned/OpenEnded\_Question:** independent reviewers 1 & 2 manually-coded open- text data of from the October 2020 survey.
        + **Data/Cleaned/Final\_Data\_June2022.csv:** clean data from the survey conducted in June 2022.
     3. *Data/Auxiliary.* This folder contains data on the 2019 U.S Current Population Survey (CPS) for the **subsample** of individuals who are 16 years old or older obtained through IPUMS.
     4. Detail and Data Availability Statement of Each Data Source
        + **2019 U.S Current Population Survey (CPS) for the subsample of individuals who are 16 years old obtained through IPUMS (*Data/Auxiliary/CPS\_2019.dta*).** IPUMS does not allow for redistribution, except for a subset of data for the purpose of replication archives as per <https://www.ipums.org/about/terms>. Contact IPUMS for permission for any other redistribution.
        + **The experimental survey data used in the paper contained in the folders *Data/Raw/ and Data/Cleaned/*** werecollected by the authors and are available under a Creative Commons Non-commercial license. The survey was designed and implemented to gather information on various economic behaviors and preferences from a sample of participants. The details of the survey design, sampling methodology, and data collection process are described in the main text of the paper. The survey questions used to collect the data can be found in the paper's appendix. For any further information about the data or regarding its use, please contact the corresponding author of the paper.

1. Description of Programs/Code

There are two main files which run scripts in the *Code* folder:

* 1. **Master Stata.do.** code runs the scripts do\_datacleaning.do, and do\_main.do. Each script generates the following output:
     + **do\_datacleaning.do:** cleans the data contained in the folder *Data/Raw* and place the clean data used for the analysis in the folder *Data/Clean*.
     + **do\_main.do:** generates Figures 2 and 3, and Tables 1, 3, 4, 5, 6, 7, 8, 9, and 11.
  2. **Master R.R.** code runs the script R\_main.R. Each script generates the following output:
     + **R\_main.R:** generates Figure 4, and Table 10.

1. Computational Requirements

The full code can be run using Stata/MP 15.1 and R 4.0.3. The code was developed and tested on a Mac with Mac computer with 16GB of RAM.

The following are the non-default Stata packages that are needed for the programs to run, together with instructions on how to install them:

* + - **distinct**: to install the latest version write *”ssc install distinct”* in the command window in Stata.
    - **outreg2**: to install the latest version write *”ssc install outreg2”* in the command window in Stata.
    - **parmest**: to install the latest version write *”ssc install parmest”* in the command window in Stata.

The list of R packages needed to run the R codes are already on the scripts. The full code can be run in less than 10 minutes on most machines.

1. Instructions for Replications
   * Make sure to change the directory to the main Replication folder on Master\_Stata.do and Master\_R.R.
   * Run Master\_Stata.do and Master\_R.R to create the datasets and generate the figures and tables in the same sequence as they appear on the paper.

# Bibliography

Flood, S., M. King, R. Rodgers, S. Ruggles, and J. R. Warren (2022). Integrated public use microdata series, current population survey: Version 9.0 [dataset]. <https://doi.org/10.18128/D030.V9.0>. March 2019 CPS data.