Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

# Part 1: Data

☐ This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

## Abstract

We have two experiment datasets.

The ``opportunity knocks" (OK) experiment was a stratified randomized experiment launched to evaluate the effect of financial incentive on college students' academic performance. The dataset contains the applicants’ background information such as high school grade, gender as the pretreatment covariates and the grades at the end of the semester as the outcome.

In the experiment of Lee (2022), 1006 Americans over 18 years old were recruited from an online survey panel. The survey contains the demographic question such as age, gender, ethnic group, edcation type. The outcome is an index ranging from 0 to 1, with higher values indicating higher generalized social trust. Some population-level measures is download from the website of the United states Census Bureau.

## Availability

☒ Data **are** publicly available

☐ Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available dat*a section, below.

### Publicly available data

☐ Data are available online at:

☐ Data are available as part of the paper’s supplementary material.

☐ Data are publicly available by request, following the process described here:

☒ Data are or will be made available through some other mechanism, described here:

Download OK dataset at: <https://economics.mit.edu/people/faculty/josh-angrist/angrist-data-archive>.

Download ``Social Trust” dataset at: <https://doi.org/10.7910/DVN/WXOYVM>.

Download population summary measure of age and sex (``nc-est2020-agesex-res”) from <https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/state/asrh/>.

Download population summary measure of race (`` SC-EST2020-ALLDATA5.csv”) from https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/national/asrh/.

### Non-publicly available data

Discussion of lack of publicly available data:

## Description

### File format(s)

☒ CSV or other plain text:

☒ Software-specific binary format (.Rda, Python pickle, etc.): .dta

☐ Standardized binary format (e.g., netCDF, HDF5, etc.):

☐ Other (described here):

### Data dictionary

☐ Provided by the authors in the following file(s):

☒ Data file(s) is (are) self-describiing (e.g., netCDF files)

☐ Available at the following URL:

### Additional information (optional)

# Part 2: Code

## Abstract

The code implements various methods in the article, generates artificial data sets, preprocesses real data sets, and outputs various figures in the paper.

## Description

### Code format(s)

☒ Script files

☒ R ☐ Python ☐ Matlab

☐ Other:

☐ Package

☐ R ☐ Python ☐ MATLAB toolbox

☐ Other:

☐ Reproducible report

☐ R Markdown ☐ Jupyter notebook

☐ Other:

☐ Shell script

☐ Other (described here):

### **Supporting software requirements**

Version of primary software used

R version 4.3.1

Libraries and dependencies used by the code

MASS 7.3.60,

tidyverse 2.0.0,

doParallel 1.0.17,

sandwich 3.0.2,

glmnet 4.1.8,

leaps 3.1,

ggplot2 3.4.3,

gridExtra 2.3,

formula.tools 1.7.1

### Supporting system/hardware requirements (optional)

### Parallelization used

☐ No parallel code used

☒ Multi-core parallelization on a single machine/node

Number of cores used:  48

☐ Multi-machine/multi-node parallelization

Number of nodes and cores used:

### License

☒ MIT License (default)

☐ BSD

☐ GPL v3.0

☐ Creative Commons

☐ Other (described here):

### Additional information (optional)

# Part 3: Reproducibility workflow

## Scope

The provided workflow reproduces:

☒ Any numbers provided in text in the paper

☒ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

☒ All tables and figures in the paper

☐ Selected tables and figures in the paper, as explained and justified here:

## Workflow details

### Format(s)

☐ Single master code file

☐ Wrapper (shell) script(s)

☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

☒ Text file (e.g., a readme-style file) that documents workflow

☐ Makefile

☐ Other (more detail in 'Instructions' below)

### Instructions

Details are in the README.txt. Small-scale simulations do not require parallel computing. One can change the setup variable to specify the scale of the simulation.

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

☐ <1 minute

☐ 1-10 minutes

☐ 10-60 minutes

☐ 1-8 hours

☒ >8 hours

☐ Not feasible to run on a desktop machine, as described here:

### Additional documentation (optional)

# Notes (optional)