

# Replication package for the paper “Marginal Effects for Probit and Tobit with Endogeneity”

## README

### Overview

The code of this replication package runs the simulation study and implements the empirical analysis using Matlab. The main program calls all the necessary programs to generate Figure 1 and Tables 1-3 in the paper.

To replicate the figure and the tables please run the program `RUN_ALL.m`.

### Data Availability and Provenance Statements

The paper uses public-use data from NLSY97 survey.

The NLSY97 survey is sponsored and directed by the U.S. Bureau of Labor Statistics and managed by the Center for Human Resource Research (CHRR) at The Ohio State University. Interviews are conducted by the National Opinion Research Center (NORC) at the University of Chicago.

The data was obtained from <https://www.nlsinfo.org/investigator>. The raw data is in the directory `NLSY/nlsy97_raw/`. The empirical analysis uses the dataset `NLSY/nlsy97_empirical.csv`. File `NLSY/nlsy97_empirical.csv` can be generated from the raw data by running `make_dataset.m`.

### Statement about Rights

The authors of the manuscript certify that they have legitimate access to and permission to use the data used in this manuscript.

### Summary of Availability

All data are publicly available.

### Dataset list

Data File	Source	Notes	Provided
<code>NLSY/nlsy97_empirical.csv</code>	NLSY97	Main dataset for the empirical analysis	TRUE
<code>NLSY/nlsy97_raw/nlsy97_v0.csv</code>	NLSY97	Raw data, used to produce <code>nlsy97_empirical.csv</code>	TRUE

### Software used

#### Computational requirements

##### Software Requirements

Matlab (the code was last run with Matlab Release 2024b for Windows)

Name	Version	Release
-----	-----	-----
{'MATLAB'}	{'24.2'}	2024b
{'Optimization Toolbox'}	{'24.2'}	2024b
{'Statistics and Machine Learning Toolbox'}	{'24.2'}	2024b
{'Parallel Computing Toolbox'}	{'24.2'}	2024b

### Controlled Randomness

Random seed is set at line 66 of `MC_replicate_main.m`.

## Memory, Runtime, Storage Requirements

**Summary** Approximate time needed to reproduce the analyses on a standard 2024 desktop machine: 10-60 minutes

Approximate storage space needed: 250 MB - 2 GB

**Details** The code was last run on an Intel i7-12700H laptop with 32Gb of RAM, using Matlab 2024b on Windows 11.

## Description of programs/code

Single main program `RUN_ALL.m` calls all the required programs. The running times are given for an Intel i7-12700H laptop with 32Gb of RAM, using Matlab 2024b on Windows 11. The main program `RUN_ALL.m`

- calls `MC_replicate_main.m`, which runs the simulations and replicates Figure 1 of the paper as Matlab figure. (Running time is about 15 minutes. Produces an intermediate 150Mb mat file `fig_MC_R5000__n1000.mat`.)
- calls `xobit_empirical_application.m` for the Tobit model, which replicates Tables 1-2 of the paper. The two Tables appear both in the Matlab command window and `Empirical_res/diary_Tobit_NLSY97.txt`. (Running time < 1 minute.)
- calls `xobit_empirical_application.m` for the Probit model, which replicates Table 3 of the paper. The Table appears both in the Matlab command window and `Empirical_res/diary_Probit_NLSY97.txt`. (Running time < 1 minute.)
- Also, optionally, if the corresponding lines are uncommented, the main program calls:
  - it calls `make_dataset.m`, which cleans the raw data in `NLSY/nlsy97_raw/nlsy97_v0.csv` and produces the dataset `NLSY/nlsy97_empirical.csv`. (Running time < 2 minutes.)
  - it calls `make_stored_grid_CV_for_max.m` that recreates file `stored_grid_CV_for_max_1-sided.mat`, which contains the quantiles of the bivariate standard Normal distribution used by the program `inference_on_max`. (Running time is about 20 minutes.)

## Instructions to Replicators

Please run the program `RUN_ALL.m`.

## References

Bureau of Labor Statistics, US Department of Labor (2024). National longitudinal survey of youth 1997 cohort, 1997–2021 (rounds 1–20). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH.

---

## Acknowledgements

This file was prepared based on the [https://github.com/social-science-data-editors/template\\_README](https://github.com/social-science-data-editors/template_README).