Guide to the data and codes used in "Labor Market Participation, Marriage and Individual Welfare"*

Laurens Cherchye, Bram De Rock and Khushboo Surana

Overview

All the results presented in the paper and the supplementary online appendix can be replicated using the files in the folders `PSID`, `Main` and `Appendix`. With the exception of the tables in Appendix D.1, D.2, D.3, D.4 and D.5, all the tables and figures can be obtained using the files in the folders `PSID` and `Main`. Rest of the tables can be obtained using the files in folder `Appendix`.

Data

Our dataset is drawn from a sample of U.S.-based household data from the 2019 wave of the Panel Study of Income Dynamics (PSID). The PSID data are produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI. They cannot be redistributed under any circumstances but can be downloaded from https://psidonline.isr.umich.edu/ upon registration and acceptance of the terms of use. Access is free of charge. On the PSID

*The Viking cluster was used during this project, which is a high performance compute facility provided by the University of York. We are grateful for computational support from the University of York, IT Services and the Research IT team.

website, the data source we use is listed under "Family Files" in the PSID's "Main Study" category. Only data from 2019 are needed for replication.

The PSID allows researchers to safely store data extracts, code, and documentation for research projects through the ICPSR's Data Repository. For replication purposes, we have uploaded our selected sample data and all results to the Panel Study of Income Dynamics Repository project openicpsr-219001. Replicators can download all necessary files for replication from https://www.openicpsr.org/openicpsr/project/219001/version/V1/view.

Code

Software

The programs for the paper are written in STATA (version 18.0) and MATLAB (version R2022a). The optimization programs in MATLAB use YALMIP (see Lofberg, 2004) and the MOSEK optimization solver. YALMIP can be downloaded for free from https://yalmip.github.io/ (copyright owned by Johan Löfberg).

After downloading YALMIP, the replicator needs to add its directories using the function 'addpath'. See https://yalmip.github.io/tutorial/installation/ for more information. The optimization problems have been solved using the MOSEK solver. An academic version of MOSEK is available for free from https://www.mosek.com/products/academic-licenses/. For installation instructions, see https://docs.mosek.com/latest/install/installation.html.

Instructions for Replication

The empirical findings can be replicated as follows.

1. Download the replication package and the data package from the ICPSR's Data Repository. Both packages follow the same naming structure. After unzipping both packages, place the files from each folder in the data package into the corresponding folder in the replication package.

2. Open `PSID/Data_cleaning.do` in Stata, specify the root directory of the raw 2019 family file, and run the program. This step only takes a few seconds and returns `PSID2019.dta` and `PSID2019.xlsx` as output files.

3. Open MATLAB, import `PSID2019.xlsx`, rename it, and save it as `Data.mat`. This file will serve as input for all identification exercises.

4. The following folders contain the necessary code to replicate the tables presented in the paper. Each folder includes a read-me file with further details.

   - Folder `PSID` replicates Tables 1-5, 13, 17-22 and parts of Tables 7-12.
   - Folder `Main/RICEB_4Type` replicates Tables 6, 15 and 16.
   - Folder `Main/RICEB_16Type` replicates Tables 7, 8 and 9.
   - Folder `Main/Povrate_5Type` replicates Table 10.
   - Folder `Main/Povrate_16Type` replicates Tables 11 and 12.
   - Folder `Main/CEB_individual` replicates Table 14 and Figures 1-9.
   - Folder `Appendix/VoluntaryInvoluntary` replicates Table 23.
   - Folder `Appendix/FulltimeParttime` replicates Table 24.
   - Folder `Appendix/BoundingWages` replicates Tables 25, 26 and 27.
   - Folder `Appendix/BartenScales` replicates Tables 29 and 30.
   - Folder `Appendix/SubsampleSize` replicates Tables 31 and 32.
   - Folder `Appendix/RandomSubsamples` replicates Table 33.

   Note: The identification steps are computationally intensive and may take several hours for each estimation. However, in the first lines of the "main" MATLAB codes, the replicator can reduce the number of iterations to decrease computing time. We used the Viking Cluster, a high-performance computing facility provided by the University of York, to parallelize the code and reduce overall runtime. If the replicator wants to run the programs on a local machine, they can set "cluster = 0" in the first lines of the code. They may also use a

different solver (e.g. GUROBI) and adjust the "parnum" parameter to optimize the program for their specific machine. With these adjustments, the empirical results may differ from those in the paper, but this will allow the replicator to verify that the program runs correctly.

5. We have included the result files from our computations in the data package. All empirical results can also be verified by directly uploading the `.mat` file into MATLAB (see the "Read me" files) and printing the results (using the last few lines of the "main" MATLAB codes).

References

J. Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*, pages 284–289. IEEE, 2004.