README

Overview

The code in this replication package constructs the analysis file from the Delgado, Garcia-Suaza and Sant'Anna (2021) using R and Stata. The simulation results for the Monte Carlo simulations are generated using the `simulation/KMDR-sim-main.R` file, whereas the `application/run-all.R` generates all the results for the empirical application.

Data Availability and Provenance Statements

Statement about Rights

- ☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☑ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

Summary of Availability

- ☑ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Details on Data Source

This paper uses data from Survey of Income and Program Participation (SIPP) for the period spanning 1985-2000, analyzed by Chetty (2008). Data can be downloaded from https://rajchetty.com/wp-content/uploads/2021/04/Chetty_UI_stata_code.zip. A copy of the raw data and Chetty's replication files is provided is provided as part of this archive at `data/Chetty2008/SIPP`. A copy of the processed (merged) data is available at `data/raw`.

Datafile: `data/raw/SIPP_durations.dta`

Computational requirements

Software Requirements

- Stata (code was last run with version 17)
  - the program "`application/00-data-process.do`" will create the data we use in the application. Make sure to set up the absolute path of the `data/raw` folder in line 10.
- R 4.1.0
  - `here` 1.0.1
  - `haven` 2.4.3
  - `survival` 3.2-11

- `matrixStats` 0.61.0
- `ggplot2` 3.3.5
- `gridExtra` 2.3
- `ggthemes` 4.2.4
- `latex2exp` 0.9.0
- `glue` 1.4.2
- `RColorBrewer` 1.1-2
- `patchwork` 1.1.1
- `utils` 4.1.0
- `foreach` 1.5.1
- `doRNG` 1.8.2
- `doSNOW` 1.0.19
- `icenReg` 2.0.15
- `TransModel` 2.1
- `readr` 2.1.0
- `taRifx` 1.0.6.2

All programs were run on a Windows 11 machine.

Controlled Randomness

- ☑ Random seed for the Monte Carlo simulations is set at line 13 of program `simulations/KMDR-sim-main.R`
- ☑ Random seed for the application is set at line 15 of program `application/01-R-prep.R`

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard 2021 desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☑ 1-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Details The code was last run on a **24-core Intel-based Desktop with Windows 11 and 32 GB of RAM**. Computation for the empirical application took 3 hours and 22 minutes. Computation for the Monte Carlo simulations took 58 minutes.

Description of programs/code

- Programs in `application` will extract the dataset and generate all results for the empirical application. The file `application/run-all.R` will run all the `R` programs.
  - The `Stata` file `application/00-data-process.do` generate variables and labels for the variables we use in the application. It saves the processed data, `data//processed/sipp-processed.dta`.
  - The `R` file `application/01-R-prep.R` prepare data for the replication, and load all functions into environment. It also generates the relevant subsets of the data for the analysis.
  - The `R` file `application/02-pooled.R` estimates the Kaplan-Meier distribution regressions and the Cox-Proportional hazard model using the pooled dataset.
  - The `R` file `application/03-pooled.R` estimates the Kaplan-Meier distribution regressions and the Cox-Proportional hazard model using the subsets of the data depending on whether households are above or below the net liquid wealth median.
  - The `R` file `application/04-mortgage.R` estimates the Kaplan-Meier distribution regressions and the Cox-Proportional hazard model using the subsets of the data depending on whether households have a mortgage or not.
  - The `R` file `application/05-hyp_tests.R` test whether the Kaplan-Meier distribution regression coefficients are constant across elapsed duration. This is the slowest portion of the code, which takes around 3 hours to run. If not interested, please comment it out.
  - The `R` file `application/06-plots.R` generates figure 1 (`figures/ub-adme-baseline.pdf`), figure 2 (`figures/ub-adme-liquidity.pdf`), and figure 3 (`figures/ub-adme-het-liquidity.pdf`) of the paper.

- Programs in `simulations` will generate all results for the Monte Carlo simulations. The file `simulations/KMDR-sim-main.R` will run all the `R` programs and generate all results.
  - The folder `simulations/aux_functions` contain auxiliary functions called by the file `KMDR-sim-main.R`. It contains the function that generate the DGP (`dgps-sim.R`), the script to compute all computations in each Monte Carlo setup (`simulations.R`), and the files that script that summarize all results (`table_dgp1.R`, `table_dgp2.R`, and `table_dgp3.R`). The file that generate Table 1 in the paper is `table_dgp3.R`; this table coincides with Table 4 in the Online Appendix.
  - The folder `simulations/results` contain all `csv` outputs generated by the simulations.
  - The folder `simulations/tables` contain summaries of the Monte Carlo results. The results for Table 1 in the paper can be found in

`table_dgp3_summary.txt`; this table coincides with Table 4 in the Online Appendix.

Instructions to Replicators

- Edit line 10 of the `Stata` do-file "`application/00-data-process.do`" to adjust the default absolute path of the `data/raw`.

- To replicate the empirical application results, we recommend you to have `R Studio` installed in your computer. Once that is installed, open the `R project` file `KMDR.Rproj`, and then run the program `application/run-all.R` to run all steps in sequence.

- To replicate the Monte Carlo simulations results, open the `R project` file `KMDR.Rproj` (if that is not already open), and then run the program `application/run-all.R` to run all steps in sequence.

List of tables and programs

The provided code reproduces:

☑ All numbers provided in text in the paper
☑ All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below.

Figure/Table #	Program	Line Number	Output File
Table 1	`simulations/KMDR-sim-main.R`	76	`simulations/tables/table_dgp3_summary.txt`
Table 2	`simulations/KMDR-sim-main.R`	74	`simulations/tables/table_dgp1_summary.txt`
Table 3	`simulations/KMDR-sim-main.R`	75	`simulations/tables/table_dgp2_summary.txt`
Table 4	`simulations/KMDR-sim-main.R`	76	`simulations/tables/table_dgp3_summary.txt`
Figure 1	`application/06-plots.R`	369	`figures/ub-adme-baseline.pdf`
Figure 2	`application/06-plots.R`	378	`figures/ub-adme-liquidity.pdf`
Figure 3	`application/06-plots.R`	386	`figures/ub-adme-het-liquidity.pdf`

References

U.S. Census Bureau. Survey of Income and Program Participation (SIPP), 1985-2000. https://www.census.gov/programs-surveys/sipp.html.

Chetty, R. (2008). Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of Political Economy* 116, 173–234.

Chetty, R. (2008). Moral Hazard versus Liquidity and Optimal Unemployment Insurance - Replication package. *Journal of Political Economy* 116, 173–234. Available at https://rajchetty.com/wp-content/uploads/2021/04/Chetty_UI_stata_code.zip (last accessed: Jan 26 2022).

Delgado,M., García-Suaza, A. and Sant'Anna, P. H. C. (2022). Distribution Regression in Duration Analysis: an Application to Unemployment Spells. *Econometrics Journal* Forthcoming.