Readme for the replication of "Household Self-Insurance and Selection into Disability Insurance"

Journal of Political Economy, 2025

Maxwell Kellogg

The replication code for this paper is broken down into three parts. The project uses Stata and R scripts in different places. Necessary packages are loaded in the 'master' R scripts.

Data Cleaning for the PSID is done in stata do-files A1 through A5, found in the root folder. These produce a core analytic file (**Full_Panel_2b.dta**) and files used for the event study analyses (**/eventstudy/psid_cleaned_headansyp.dta**). The analytic files used for the SIPP and the HRS are constructed by **sipp_setup.do** and **hrs_setup.do** respectively, but the raw data files that they reference are not included in this package due to their sizes. The raw HRS file is the RAND HRS longitudinal file, and the raw SIPP files can be downloaded from the SIPP home page. The analytic dataset files for all three datasets have been included in this replication package. I also provide three files derived from CES data (**b0.dta**, **finprice.dta**, **natpr.dta**) which were used in imputation methods that did not appear in the final paper, but which may be exist in certain parts of the code.

Estimating basic empirical moments for the structural model is done using PSID data in stata do-file **B0_makemoments.do**, and the do-files in the /makemoments/ directory which it references. Moment estimates have been included in the file, so that one should not need to rerun this segment of the project in order to estimate the model.

Estimation of event study results is done in the R script **C1_master_evenstudy.R**, and the R scripts that it references in the /eventstudy/ directory. Some of these estimates are referenced by the model, so estimates have been included in the file so that one should not need to rerun this segment of the project in order to estimate the model.

The structural model is estimated and studied in R scripts D1 through D8. The model and standard errors associated with it were estimated using an HPC. The batch script, **batchscript.sh** is provided; I used a cluster of 6 nodes each running in parallel with 10 threads. The model was coded in C++ and is provided as an R package "spousalVFIParallel" which is the **spousalVFIParallel_1.4.tar.gz** file.

- **D1_master_model.R** loads auxiliary functions and calls all the remaining scripts.
- **D2_Solve_ParallelSpouseC_pso_fixmar.R** estimates the structural model. The output from my last call of this script is provided in the /modeloutput/particlefiles directory, so that one can study the model without re-estimating it (the most computationally intensive step).
- **D3_Solve_Varcov.R** estimates the asymptotic variance of model parameters, producing **newguess_varcov_simplified.csv** (the variance-covariance matrix) and **newguess_sensitivity_simplified.csv** (the sensitivity matrix).

- **D4_Solve_Varcov_Bootstrap.R** estimates the variance of other features derived from the model, based on the asymptotic distribution of the model parameters. Output has been provided in the/modeloutput/bootstrap directory.
- **D5_Welfare_wtpdecomp.R** uses the fitted model to do welfare analysis, producing a RDS file (**wtp_decompose.RDS**) which stores them, to be organized and printed in script D8.
- **D6_kmean_simulations.R** does sensitivity analysis on the type classification procedure (Table A.5).
- **D7_Welfare_evalfit2.R** assesses the fit of the model, producing Figure 3, Table A.3, and Table A.4.
- **D8_Plot_wtp_decompse2.R** uses the output of the previous R scripts to organize and print most of the model-related tables and figures (Table 3, 4, 5, 6, A.5, A.6, A.7, A.8, H.1, and I.1 and Figure 4). Note, though, that Figure A4 is generated in Stata do-file **model_fitting.do** which is called instead by the do-file **B0_makemoments.do.**