

Replication Package for “Endogenous Information and Simplifying Insurance Choice”

README

The code in this replication package constructs analysis files from Medicare claims and enrollment data using Stata. The Medicare claims data are restricted use and cannot be shared, however we provide code to generate a synthetic analysis dataset. Stata and MATLAB are used to create simulations from the theoretical model, run the descriptive analysis, run the structural model, and output estimates and counterfactual results.

Data Availability Statement

The Medicare data for this project are confidential, but may be obtained with Data Use Agreements with the Centers for Medicare and Medicaid Services. Researchers interested in access to the data may do so through ResDAC by visiting <https://resdac.org/> or contacting resdac@umn.edu. Alternatively, NBER affiliates may access the data through NBER (<https://www.nber.org/research/nber-internal-data-collection>). It can take some months to negotiate data use agreements and gain access to the data. The author will assist with any reasonable replication attempts for two years following publication.

Details on each Data Source

There are three Medicare files used to construct the analysis dataset: raw claims, plan characteristics, and formulary files (CMS 2010-2015). The construction of the analysis dataset follows Abaluck and Gruber (2011) and Abaluck and Gruber (2016) closely, however the code is updated to use more recent Medicare claims data. For a detailed description of the variables see the replications packages (<http://doi.org/10.3886/E112428V1> and <http://doi.org/10.3886/E112927V1>).

We also use publicly available data on plan star ratings and zip code characteristics. Plan star ratings come from the Centers for Medicare and Medicaid Services (<https://www.cms.gov/medicare/health-drug-plans/part-c-d-performance-data>). Zip code characteristics come from the American Community Survey (<https://www.census.gov/programs-surveys/acs/data.html>).

Software Requirements

Stata (code was last run with version 17)

- Install Stata packages using `install_packages.do` (all packages installed as of 2018-07-14)
- Also install `guse` and `gsave` from <https://sites.google.com/site/johannesschmieder/stata>

Matlab (code was run with Matlab Release 2018b)

- Econometrics package
- optimization package
- Multiprecision Computing Toolbox (AdvantixMCT-4.6.0.13135). Not necessary to run code but allows for standard errors from the structural model that are more accurate.

Portions of the code use bash scripting, which may require Linux. When using the bash script (run_all_with_simulated_data_pc.sh) on Windows rather than Linux, the following code,

```
matlab -nosplash -nodesktop < sim_stdoop_welfare.m > sim_stdoop_welfare.log
```

should be modified to:

```
matlab -nosplash -nodesktop -logfile sim_stdoop_welfare.txt -r " sim_stdoop_welfare; exit"
```

Runtime Requirements

Programs run on servers with 32 cores and 128GB of ram. Running the code to build the data takes about 30 hours. Running the analysis code takes 25 hours (<15 minutes with the synthetic data).

Description of programs/code

- Programs in simulation/ folder create charts using a simulation from the theoretical model.
 - simulation/sim_stdoop_welfare.m simulates data from the theoretical model
 - simulation/make_charts.do makes charts using simulated data from the theoretical model
- Programs in build/ folder import the data and construct out-of-pocket costs for each plan in an individual's choice set following Abaluck and Gruber (2011) and Abaluck and Gruber (2016). Code follows Abaluck and Gruber (2016) closely with updates to use more recent data.
 - simulated_data/make_simulated_data.do makes synthetic version of the main analysis datasets
- Programs in descriptive/ folder create summary statistics and charts and figures showing motivating evidence.
 - See list of tables and programs below for additional details.
- Programs in struct/ folder estimate the structural model and simulate counterfactuals.
 - struct/mainspec/run_estimation.m runs the baseline structural model
 - struct/mainspec/*.do files examine model fit and counterfactuals. See list of tables and programs below for additional details.
 - struct/output_main_est_for_latex.m outputs the results for latex
 - The folders for alternative specifications are structured similarly

- Programs in montecarlo/ folder examine the sensitivity of the model to distributional assumptions using simulated data for Appendix G.

License for Code

The code is licensed under a BSD license.

Instructions to Replicators

In order to run the replication using synthetic data, the sequence of programs can be run from the bash shell script file `run_all_with_simulated_data_mac.sh` or `run_all_with_simulated_data_pc.bat` in the main directory (change the default path at the top of the program). The synthetic data has similar broad patterns as the actual data. Using the synthetic data, the main specification estimates a mean marginal cost of information of 4, mean price sensitivity of -1.9, and an average total information acquisition cost of 263.

To replicate the analysis with Medicare data, first see `build/BUILD_README.md`. After running code to build the analysis datasets, replace the synthetic data and run the programs listed in `run_all_with_simulated_data.sh`.

List of tables and programs

List of programs corresponding to the main article tables and figures, with the relevant script:

- Figure 1: `simulation/make_charts.do`
 - Outputs `model_overspending_by_stakes_chart.pdf` and `model_logit_coef_by_stakes_chart.pdf`
- Table 1: `descriptive/summary_stats_all.do`
 - Outputs `summary_stats_a.tex` - `summary_stats_f.tex`
- Figure 2: `descriptive/overspending_by_stakes_chart.do`
 - Outputs `overspending_by_stakes_chart_simple.pdf`
- Table 2: `descriptive/overspending_by_stakes_regs.do`
 - Outputs `overspending_by_stakes_regs_bins_a.tex` and `overspending_by_stakes_regs_bins_b.tex`
- Figure 3: `descriptive/logit_coef_chart.do`
 - Outputs `logit_coef_chart.pdf`
- Table 3: `descriptive/logit_coef_stakes_interaction.do`
 - Outputs `logit_coef_interaction_a.tex` and `logit_coef_interaction_b.tex`
- Table 4: After running `mainspec/run_estimation.m`, results are outputted by `output_main_est_for_latex.m`

- Outputs demand_estimates.tex
- Table 5: struct/mainspec/s02_sim_summary_stats.do
 - Outputs cf_table_a.tex - cf_table_d.tex
- Figure 4: struct/mainspec/s07_sim_restricted_choice.do
 - Outputs rest_age_plan_choiceset_welfare_chart.pdf and rest_age_plan_choiceset_welfare_alt_chart.pdf
- Table 6: Results are outputted by struct/mainspec/output_counterfactual_summary.m
 - Outputs cf_table2_a.tex
- Table 7: After running struct/heteropriorspec/run_estimation.m, struct/cell1/run_estimation.m, and struct/cell2/run_estimation.m, results are outputted by output_all_heteroprior_for_latex_w_cf.m
 - Outputs demand_estimates_cell_w_cf.tex

List of programs corresponding to the Online Appendix tables and figures.

- Table A-1: struct/mainspec/s04_sim_logit_chart_alt_models.do
 - Outputs alt_model_est_a.tex and alt_model_est_b.tex
- Table A-2: struct/mainspec/output_counterfactual_summary_alt_models.m
 - Outputs cf_table2_alt_models_a.tex, cf_table2_alt_models_b.tex, and cf_table2_alt_models_c.tex
- Figure A-1a: struct/mainspec/s05_sim_overspending_by_stakes_chart.do
 - Outputs overspending_by_stakes_chart.pdf
- Figure A-1b: struct/mainspec/s03_sim_logit_chart.do
 - Outputs sim_logit_coef_chart_baseline.pdf
- Figure A-1c: struct/mainspec/s06_sim_overspending_by_stakes_chart_model_comparison.do
 - Outputs overspending_by_stakes_chart_model_comparison_werror.pdf
- Figure A-1d: struct/mainspec/s04_sim_logit_chart_alt_models.do
 - Outputs sim_logit_coef_chart_model_comparison_werror.pdf
- Table A-3: descriptive/overspending_by_stakes_regs_pf.do
 - Outputs overspending_by_stakes_regs_bins_pf_a.tex and overspending_by_stakes_regs_bins_pf_b.tex
- Table A-4: descriptive/overspending_by_stakes_regs_first_time_enrollees.do
 - Outputs overspending_by_stakes_firsttime_regs_a.tex and overspending_by_stakes_firsttime_regs_b.tex
- Table A-5: descriptive/logit_coef_stakes_interaction_pf.do
 - Outputs logit_coef_interaction_pf_a.tex and logit_coef_interaction_pf_b.tex
- Figure A-2: descriptive/alt_choice_quality_by_stakes_chart.do
 - Outputs chose_ltp10_by_stakes_chart.pdf and chose_ltp20_by_stakes_chart.pdf
- Figure A-3a: descriptive/chosen_risk_by_stakes_chart.do
 - Outputs chosen_lowestrisk_by_stakes_chart.pdf
- Figure A-3b: descriptive/chosen_quality_by_stakes_chart.do

- Outputs chosen_starquality_by_stakes_chart.pdf
- Table A-6: After running struct/nofe/run_estimation.m and struct/woutliers/run_estimation.m, results are outputted by struct/output_appendix_for_latex.m
 - Outputs demand_estimates_robustness.tex
- Figure A-4: struct/mainspec/s07_sim_restricted_choice.do
 - Outputs restricted_choiceset_welfare_chart.pdf and restricted_choiceset_welfare_alt_chart.pdf
- Figure A-5: struct/mainspec/s08_sim_oop_cap.do
 - Outputs oopcap_welfare_chart_all.pdf
- Figure A-6: struct/mainspec/s08_sim_oop_cap.do
 - Outputs oopcapadj_welfare_chart_all.pdf
- Table A-8: montecarlo/montecarlotable.m
 - Outputs montecarlos.tex

References

Abaluck, Jason and Jonathan Gruber, “Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program,” *American Economic Review*, 2011, 101 (4), 1180–1210.

Abaluck, Jason and Jonathan Gruber, “Evolving choice inconsistencies in choice of prescription drug insurance,” *American Economic Review*, 2016, 106 (8), 2145–84.

Centers for Medicare and Medicaid Services, 2010-2015. “Prescription Drug Event Files, Beneficiary Summary Files, Characteristics Files, and Formulary Files,” Available through Research Data Assistance Center (ResDac) with file descriptions at <https://resdac.org/cms-data>.