Replication package for Dickstein, Ho, and Mark (2023)

The Structure of the Data Archive

The data archive consists of 4 folders: data, data_construction, library, and analysis.

The original data lives in the folder 'data/orig'. The programs in the folder 'data_construction' transform those data into datasets that can be used in the estimation and counterfactual procedures. The folder 'library' contains code that is used often in the programs. The folder 'analysis' contains code that estimates the demand and premium setting models, simulates the counterfactuals, and creates the figures and tables presented in the paper.

Below, we describe the data provided in this data archive and the program package.

Data

The following datasets are included in the data archive in the 'data/orig' folder:

- *carrier_collapsed_mlr_allstates.dta, carrier_collapsed_mlr_indiv.dta, carrier_collapsed_mlr_oregon.csv,* and *cleaned_mlr_allstates.dta.* These files contain information derived from the publicly available MLR datasets.
- The *CHIS* folder contains data from the California Health Interview Survey available for download by researchers at https://healthpolicy.ucla.edu/chis/data/pages/getchisdata.aspx.
- *UninsuredRateByRAYearandAge.txt.* This file contains an estimate of the uninsurance rate for each rating area and each year from 2012 to 2017 for certain age bins. These data were derived from the *2014-2016 American Community Survey 1-year Estimates Health Insurance Coverage Status by Sex by Age* (ACS) datasets from the U.S. Census Bureau. The process by which this dataset is constructed is described in detail Online Appendix Section B.4.
- The *Risk Adjust, Reins, Risk Corridors* folder contains data on risk adjustment, reinsurance, and risk corridor payments.

The following datasets are needed to run the replication package, but are not included in the package because they are derived from proprietary datasets:

- *acg_positive_quintiles.* This is a one-line comma-separated values file that holds the quintiles of the ACG scores of households in our sample.
- *ConstructedPlanDat.csv.* This dataset contains information such as premiums, copays, and deductibles of each constructed plan.
- *ChronicConditionsMed.* The files in this folder contain chronic conditions information for each patient in the Oregon claims data. The information consists of dummy variables for the existence of chronic conditions and counts of diagnoses associated with those chronic conditions. These data are used to infer the characteristics of the uninsured population as described in Online Appendix Section B.4.
- *MM2015_u.txt* and *MM2016_u.txt.* These datasets include patient-level information for all patients in the Oregon claims dataset including a patient's relation to a subscriber and their plan coverage. These datasets are used to track 'forced switchers' into markets other than the small group or individual markets.

- *Pat2PKeyCrosswalkNew.txt.* This dataset is a crosswalk from the variable PatID, which identifies a patient-insurer, to the variable personkey, which tracks persons across insurers.
- *SecondCheapestSilver.*csv. This dataset contains the constructed plan which corresponds with the second cheapest silver in each rating area and year, along with the plan's monthly premium.
- *SubData147.csv*, *SubData157.csv*, and *SubData167.csv*. These datasets consist of the household dataset described in section B.1 of the online appendix. It includes the variables that are constructed as described in section B.2 of the online appendix.

We use Oregon's proprietary All Payer All Claims (APAC) dataset, obtained through the Oregon Health Authority's Office of Health Analytics. More details on the process for requesting access to the APAC data can be found here: https://www.oregon.gov/oha/HPA/ANALYTICS/Pages/APAC-Data-Requests.aspx

Programs

All programs, apart from the bootstrap procedure, can be run using the file *replication_runme.sh*. *replication_runme.sh* will in turn run runme.sh scripts in each of the program folders: 'data_construction', 'analysis/estimationcode', 'analysis/prem_setting', 'analysis/counterfactuals', and 'analysis/tablesandfigures'. Before running these scripts, a working directory needs to be set both in the *replication_runme.sh* file and in the file *library/PreliminariesCode.R.* Each runme.sh script provides the execution order of the programs.

As stated above, the programs in the folder 'data_construction' transform those data into datasets that can be used in the estimation and counterfactual procedures. These built datasets are saved in the 'data' folder. In the 'analysis' folder, the subfolders 'estimationcode', 'prem_setting', and 'counterfactuals' use these data to estimate the demand, premium setting, and counterfactual results. The results of these estimation procedures are saved within the subfolders themselves and some output from these programs that are used in the paper's tables and figures are saved in the 'tablesandfigures' subfolder. Last, the 'tablesandfigures' subfolder contains programs that generate output used in the paper's tables and figures. The programs live in the folder 'analysis/tablesandfigures/source' and the output lives in the folder 'analysis/tablesandfigures/release'.

A description of the programs are provided below. Per our data use agreement, we omit programming code that could be used to re-identify confidential data elements.

data_construction

Files or Folder	
AC01CreateEstimationSample.R	Creates the household level dataset used in the analysis.
AC02AddingUninsuredData.R	Add uninsured households to the household/subscriber level dataset.
AC03CreatingExplodedChoiceData.R	Create a household-insurance plan option level dataset.
AC04LabelingSwitchersinExplDat.R	Label the small group switchers in the household-insurance plan option level dataset.
AC05CreatingSGExplodedData.R	Create a household-insurance plan option level dataset for the small group households.
AD01CreatingAllChoiceData.R	Creates a dataset with all health insurance plan options in each rating area and each year.

BL01CreateSwitcherSample.R	Create the small group switcher sample. This file also creates two tables that are saved in the analysis/tablesandfigures folder.
DA001CreateAllChoicePremPostMerger.R	Creates a dataset with all health insurance plan options in each rating area and each year for the post-merger counterfactual sample.
DE01CreateCounterfactualSample.R	Create the household level dataset used in the counterfactuals
1_subscribers.py	Takes the household/subscriber level dataset, reformats it for use in the uninsured sample construction process, and uses it to define the variable cut offs used in the uninsured sample construction process.
2_chis_uninsured.py	Uses the CHIS data to estimate conditional probabilities used to distribute characteristics to the uninsured households as described in appendix section B.4.
3_uninsured_rates.py	Combine the CHIS-derived conditional probabilities to the census data to distribute characteristics to the uninsured households as described in appendix section B.4.

prem_setting

File or folder	Description
predicted.r	This program collects an estimate of expected cost across households that is then used in the premium setting estimation procedure.
main.do	This program runs the premium setting regression. It calls helpers.do and paper_present.do.
paper_present.do	This file runs the regression and saves the results.
helpers.do	This file includes functions that can be used to collect the instrumental variables.
Bootstrap folder	The programs in this folder run the same premium setting procedure with randomly drawn households for the bootstrapping procedure.

estimationcode

File or folder	Description
adjustfortype.R	This script adjusts the data and creates the necessary variable names, depending on the type of estimation.
estimatedemand.R	This script collects the specification details, estimates demand, and saves the results. This script is called by the “call_estimatedemand.R” programs and in turn calls the other programs in this folder.
generatestderrors.R	This program generates standard errors for the demand estimates.
get_llh.R	This program defines the likelihood function.

parametrize_choice_data.R	This program takes parameter values and data to generate alpha, omega, psi, and beta0'Z for each household, predicted shares for each option and each household, and the derivative of the log likelihood function with respect to each parameter related to the choice phrase of the likelihood function.
parametrize_cost_data.R	This program takes parameter values and data to generate alpha and omega, the log likelihood of each observed cost, and the derivative of the log likelihood function with respect to each parameter related to the cost phrase of the likelihood function.
call_estimatedemand.R	A version of this program is in each of the specification subfolders. In this file, we define the specification details, then run the demand estimation procedure. The program also creates an "output" folder and saves the demand estimation results in that folder. This program calls the programs in the parent estimationcode folder.
Run_bootstrap.R	This program estimates demand models with randomly drawn households for the bootstrapping procedure. It is applied only to the preferred specifications.

counterfactuals

File or folder	Description
Library	This folder contains files that are used in many of the different counterfactuals.
DA02CreateCounterfactualMedExplodData.R	This code creates a household-option level dataset in the format used in the counterfactual exercises.
DA03aRunCounterfactualFunctions.R	This script includes functions that are used to generate counterfactuals and to report counterfactual comparisons. One of these functions applies the algorithm outlined in appendix section F.1.
DA03bRunCounterfactualGetdataready.R	This script collects household characteristics, choice sets, estimated demand parameters, and estimated price setting parameters. Calls DA02CreateCounterfactualMedExplodData.R.
DA03cRunCounterfactuals1table.R	This script sets the counterfactual specifications and calls the counterfactual code. Calls DA03aRunCounterfactualFunctions.R, DA03bRunCounterfactualGetdataready.R, and DA03dCounterfactualCode.R.
DA03dCounterfactualCode.R	This code runs the counterfactual algorithm and saves the results of the counterfactual.
DA04CreateSGSurplusData.R	Creates a dataset that can be used to estimate consumer surplus for the small group households.
DA05EstimateSGSurplus.do	This program estimates consumer surplus for the small group households
DB01RunFixedMarkupCounterfactuals.R	This program calls DA03cRunCounterfactuals1table.R to run a counterfactual for a 0%, 10%, and 25% fixed markup.

Loop_specs	This folder contains files that are used in the construction of figures 5, 6, and 7 in the paper. They run many counterfactuals in a loop.
Bootstrap	The programs in this folder run the counterfactual procedure with randomly drawn households for the bootstrapping procedure.

tablesandfigures

File or folder	Description
main.do	*This script opens the files in the paper_sumstat folder and runs the functions within those files. By running this script, you generate output related to figures 1, 2, and 3, and tables 1, 2, A2, A3, and A4.*
helpers_import.do	*This script contains functions that collect and clean data for use in generating figures 1, 2, and 3, and tables 1, 2, A2, A3, and A4.*
helpers.do	*This script contains functions that are used in generating figures 1, 2, and 3, and tables 1, 2, A2, A3, and A4.*
plot_mainf_costs.do	This script includes a function that generates figure 1 – "Distribution of monthly medical costs".
plot_mainf_prem.do	This script includes a function that generates figure 2 – "Base monthly premiums".
plot_mainf_markups.do	This script includes a function that generates figure 3 – "Medical markups".
tab_sum_stats.do	This script includes a function that generates CSV versions of table 1 – "Summary statistics on demographics variables" – and table 2 – "Summary statistics on insurance variables".
tab_admincosts.do	This script includes a function that generates a CSV version of table A2 – "Administrative costs".
tab_switchers_matched.do	This script includes a function that generates a CSV version of table A3 – "Demographics of the switchers".
tab_uninsured_demographics.do	This script includes a function that generates a CSV version of table A4 – "Demographics of the uninsured".
create_mlr_state_maps	This script collects MLR data and creates the tile maps presented in Figure A1.
cs_change_bytype.r	This script creates table A7, which analyzes how the mandated insurance counterfactual affects consumer surplus for subsets of the sample.
fit_analysis.r	This script creates table A8, which analyzes how well out model's predictions fit observed summary statistics for an out-of-sample population.
var_within_constructedplans.r	This script creates table A9, which analyzes variation of plan characteristics within a constructed plan.
justify_moralhazard_riskaversion.r	This script creates a dataset that is used to create figures A2 and A3.

moral_hazard_risk_aversion_graphs.do	This script creates figures A2 and A3, which show summary statistics that describe the identification of the risk aversion and moral hazard parameters.
create_heatmapdata.r	This script creates data that are used to create figures 5, 6, and 7. The script does so by running many counterfactuals in a loop, where parameter values are change across the counterfactuals.
create_heatmaps_andbuildata.r	This script creates the heat map figures 5 and 6, which analyze how counterfactual results change as the non-discretionary spending and moral hazard parameters change, and creates a dataset used to generate figure 7.
create_stataplots.do	This script creates figure 7, which analyzes how counterfactual results change as the fixed markup parameter is changed.
tab_derived_param.r	This script creates table 5, which presents the estimated derived parameters omega and psi.
create_ctft_params.r	This script creates data used to create table 4.
create_ctftparams_tab.do	This script creates table 4, which presents the estimated non-discretionary spending of households for different subsets of the population.
creating_adjmarkup_data.r	This script creates data used to generate figure 4.
create_boxplots.do	This script creates figure 4, which shows adjusted and unadjusted medical markups.
createlatextables	The code in this folder takes CSV files from the release folder and transforms them into LaTeX tables. The program main.R defines a function used to generate LaTeX code from a CSV file and runs all of the code. The other programs define functions to generate the LaTeX tables for each table in the paper.