Replication Instructions for "Dynamic Pricing Regulation and Welfare in Insurance Markets"

by Naoki Aizawa and Ami Ko

December 2024

1 Overview

This replication package contains all the necessary codes and the information for the data to replicate empirical and quantitative findings throughout the paper.

2 Data Availability Statement

This paper utilizes the following datasets for analysis. For each dataset, we provide a description and instructions on how interested researchers can access it.

- **NAIC LTCI Experience Reports**
  - **Description:** This is one of the main data sources in our paper, which details information about the long-term care insurance industry. We primarily use their Form A, Form C, Form 2, and Form 5 across various years.
  - **Availability:** This dataset cannot be posted publicly. Researchers can purchase it directly from the NAIC (`https://content.naic.org/publications`) or access it through institutional subscriptions to SNL Financial.
- **Long-Term Care Rate History Data**
  - **Description:** We use this data to obtain information on premium changes for each plan over time.
  - **Availability:** An interested researcher can obtain this data from California Department of Insurance (`https://www.insurance.ca.gov/01-consumers/105-type/95-guides/05-health/01-ltc/index.cfm`).
- **Health and Retirement Study (HRS) Data**
  - **Description:** We use it to calculate the demographic composition of the population and consumers purchasing LTCI.
  - **Availability:** This dataset cannot be posted publicly. Researchers can register and access the data after agreeing to the user terms at `https://hrs.isr.umich.edu/`.
- **American Community Survey (ACS) Data**
  - **Description:** We use it to construct a measure for the market size.
  - **Availability:** The data can be downloaded through IPUMS USA(`https://usa.ipums.org/usa/`).

3 Description of Sample Construction and Codes

3.1 Data Analysis

The description below applies to the files in the `descriptive` folder.

1. **Nationwide trend in supply.** Combine NAIC Form A and Form 2 using the product identifier. This produces year-insurer-product-level data containing nationwide experience. A plan is treated as an exit plan if it reports zero sales and the first issue year precedes the report year. Run `Figure1_2.R` in R to produce Figures 1 and 2 of the main text.

2. **Event study analysis.** Use NAIC Form C data, which are at the year-state-insurer-product-calendar duration level. To construct a sample of new sales, restrict the data to calendar duration=0. Aggregate to the year-state level by summing the number of active insurers and plans, and computing the average premium for newly sold plans. Run `MAIN_eventstudy.do` in STATA for the baseline event study analysis and reproduce the results in Figure 3 and Figure A.1. Run `MAIN_eventstudy_BJS.do` in STATA for the event study analysis following Borusyak et al. (2024) to reproduce Figure A.2.

3. **Rate increase analysis.** Run `Figure4.R` in R, which uses Rate History data, along with NAIC Form A and Form 2, to produce Figure 4 and Table A.7.

4. **Summary statistics of firms**. Run `TableA2.do` in STATA which uses NAIC Form C data to reproduce Table A.2.

5. **IV test**. Run `TableA3_IV_check.do` in STATA to test the IV validity reported in Table A.3.

6. **HRS sample**. Use the HRS data 1998-2014. Restrict to respondents who were responsive and at least 60 years old in 1998. Run `HRS_stats.do` in STATA to compute statistics used in the analysis.

3.2 Structural Estimation and Counterfactual Experiments

The description below applies to the files in the `estimation_simulation` folder.

1. **Sample construction.**

   (a) From the NAIC Form C data containing new sales, for each insurer in a given market, retain only its "dominant" product, which has the largest sales. This brings the data to the year-state-insurer level.

   (b) Restrict to insurers whose rate increase decisions can be observed in the Rate History data. These are insurers who had strictly positive enrollment during 2007-2016 in the state of California. For enrollment information, use Form C for years 2007-2008 and Form 5 for years 2009-2016.

   (c) Merge with Rate History data using plan identifiers. For plans with multiple rate increase requests, compute the cumulative rate increases.

(d) Predict state-contingent rate increases and claims using the procedure described in Online Appendix D.1 and D.2.

(e) Construct IVs used in demand estimation by computing changes in rate increases the year the state adopts RSR 2000 and, for each insurer in a given market, its mean prices in neighboring states. IV validity can be checked by following the procedure described earlier in 3.1 #5.

(f) For market share computation, use as the denominator the population information from the ACS.

2. **Description of Codes**

(a) Run `MAIN_demand.m` in the subfolder `matlab/demand_estimation_AK` to estimate the demand parameters.

(b) Run `MAIN_supply.m` in the subfolder `matlab/supply_estimation` to estimate the cost parameters. Table 1 and Table 2 of the main text as well as Table A.4 and Table A.5 on the Online Appendix are generated.

(c) Run `MAIN_cf.m` in the subfolder `matlab/counterfactual` to run counterfactuals. Counterfactual results of the main text (Figure 6, Table 3, Table 4 and Figure 7) as well as those in the Online Appendix (Table A.6 and Table A.8) are generated.

(d) Run `MAIN_cf_lapse.m` in the subfolder `matlab/counterfactual_lapse` to generate counterfactual results with endogenous lapses reported in Figure A.3 of the Online Appendix.

(e) To reproduce Figure A.4 of the Online Appendix, uncomment lines 176-177 of `MAIN_demand.m` and rerun its lines 180-204; and then rerun `MAIN_supply.m` and `MAIN_cf.m`.

3.3 Computing Specifications

All the data analysis programs are written in either R (version 4.4.1) or STATA (version 18.0). All quantitative analysis codes are written in MATLAB (version R2023b) and have been verified to run on macOS Sonoma 14.6.1 with an Apple M1 Max chip (10 cores) and Windows 10 Pro with an Intel® Xeon® CPU E5-2697 v4 @ 2.30 GHz (18 cores).