

## Overview

The code in this replication package employs data from five primary sources: (1) Covered California Enrollment Files, obtained directly by the author via a Public Record Act Request; (2) Product and Plans files, downloaded from Covered California [4]; (3) Rate Review Filings Data, downloaded from CMS.gov [3]; (4) the American Community Survey, downloaded via IPUMS [7]; (5) the Medical Expenditure Panel Survey, provided by AHRQ.gov [1].

The analysis employs [Stata](#), [Matlab](#), and [Artelys Knitro](#).

One script runs the entire analysis invoking Matlab directly from Stata. The replicator should expect the code to run for about 14 days, using 22 cores and 256 Gb of RAM.

- Stata packages (can be installed using the “`ssc install`” command): `texresults`, `moremata`, `gsample`, `mixlogit`
- Matlab toolboxes: Parallel Computing Toolbox, Statistics and Machine Learning Toolbox, Optimization Toolbox

## Data Availability and Provenance

The author of the manuscript have legitimate access to and permission to use the data used in this manuscript.

All data are publicly available except the original, raw enrollment file, named “`PRA_Tebaldi_file_2017.02.08.csv`”. This file contains individual-level identifiable information, providing age, gender, 5 digit zip-code, income information, proxies for household structure, and insurance choice, and date of choice.

The author provides public access to not-identifiable individual level enrollment files in which neither gender nor 5 digit zip-code are provided. Geographic information are replaced by the (broader) rating regions used in the analysis.

This data is now publicly accessible, users should provide a proper citation to this article (Tebaldi, *Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA*) and provide a link to the permanent DOI [10.5281/zenodo.8367909](https://doi.org/10.5281/zenodo.8367909).

## Details on each Dataset

All data needed to replicate results in the article are collected in the `~/Input` folder.

- `~/Input/ACS` contains the 5-year ACS sample downloaded via IPUMS [7] for 2013, 2014, 2015, 2016.
- `~/Input/De-Identified Data` contains “`individual-level-inside-YEAR.dta`” and “`individual-level-out-YEAR.dta`” for `YEAR=2014,2015,2016,2017`. These contain individual-level, not-identifiable records of enrollment matched to individual in the ACS using:

- **~/Scripts/DO/analysis/1-Clean-Administrative-File.do**
- **~/Scripts/DO/analysis/2-Household-Choice-Level-Data.do**
- **~/Scripts/DO/analysis/3-Merge-Outside-Option-ShareData.do**

*"individual-level-inside-YEAR.dta"* contain the following variables

VARIABLE NAME	VARIABLE DESCRIPTION
enrollment_year	Enrollment Year
hios_id_16	16-digits HIOS plan identifier
FPL	Federal Poverty Level
hh_count	Number of enrollees in the household
subsidy	Monthly Subsidy
age_factor	Total age adjustment factors for the household
children	Indicator = 1 if any children are enrolled
subsidized	Indicator = 1 if the household is subsidized
IDregion	Covered California Rating Region
plan_level	Metal Tier
issuer_name	Name of Insurance Carrier
plan_network_~e	Network (HMO/PPO/EPO)
monthlyprice	Monthly Premium
insurerID	ID for insurer
tierID	ID for metal tier
networkID	ID for network type
Y	Y=1 for chosen plan
insurer_revenue	Insurer Revenue (Premium+Subsidy)
av	Actuarial Value
Pyearly00	Annual Premium in USD/1000
buyerID	ID for buyer
planID	ID for Region-Insurer-Tier-Network
coarseage	Round Average Age of the Household
coarsefpl	Bin of Federal Poverty Level
mktID	ID for year-region-household size-income-age

*"individual-level-out-YEAR.dta"* contain the following variables

VARIABLE NAME	VARIABLE DESCRIPTION
buyerID	ID for buyer (in inside dataset)
replicaID	ID for replica of buyer when merged to ACS
buyerID_out	string(buyerID)+"R"+string(replicaID)
enrollment_year	Enrollment Year
hios_id_16	16-digits HIOS plan identifier (choise in inside dataset)
FPL	Federal Poverty Level
hh_count	Number of enrollees in the household
subsidy	Monthly Subsidy
age_factor	Total age adjustment factors for the household
children	Indicator = 1 if any children are enrolled
subsidized	Indicator = 1 if the household is subsidized
IDregion	Covered California Rating Region
plan_level	Metal Tier
issuer_name	Name of Insurance Carrier
plan_network_~e	Network (HMO/PPO/EPO)
monthlyprice	Monthly Premium
insurerID	ID for insurer
tierID	ID for metal tier
networkID	ID for network type
Y	Y=1 for chosen plan
insurer_revenue	Insurer Revenue (Premium+Subsidy)
av	Actuarial Value
Pyearly00	Annual Premium in USD/1000
planID	ID for Region-Insurer-Tier-Network
coarseage	Round Average Age of the Household
coarsefpl	Bin of Federal Poverty Level
mktID	ID for year-region-household size-income-age

- **~/Input/MEPS** contains the .dat files provided by [AHRQ.gov](http://AHRQ.gov) for the 2014, 2015, 2016, and 2017 Medical Expenditure Panel Survey [1].
- **~/Input/PRA Raw Data** is void, it is a placeholder for the original file "PRA\_Tebaldi\_file\_2017.02.08.csv" containing identifiable information. Appendix A of this readme includes the original email used by the author when requesting information from Covered California.
- **~/Input/Raw Auxiliary Data** collects:
  - **~/Input/Raw Auxiliary Data/Age Factors** providing a crosswalk from age to ACA age-adjustment factor from [CMS.gov](http://CMS.gov) [2].
  - **~/Input/Raw Auxiliary Data/Crosswalks** contains .dta files with zipcode-region-puma crosswalks generated using input from Census.gov [8] and the Dartmouth Atlas [5].
  - **~/Input/Raw Auxiliary Data/Product Prices** collects raw data on product and prices in Covered California during 2014-2017; downloaded from [Covered California](http://Covered California) [4].
  - **~/Input/Raw Auxiliary Data/Subsidy Schedule (IRS)** contains the crosswalk from income to FPL and maximum affordable amount; obtained from the [IRS Instructions for Form 8962](http://IRS Instructions for Form 8962) [6].
- **~/Input/RRF** contains Public Use Files of Rate Review Filings Data, downloaded from [CMS.gov](http://CMS.gov) for 2016, 2017, 2018, and 2019 [3].

## Description of programs/code

- **~/Scripts/DO/MASTER\_DO\_replication.do** will run the entire analysis and generate all exhibits in the article.
- This invokes, in order, the following scripts collected in **~/Scripts/DO/analysis**:
  - **"analysis/00-ACS-to-potentialbuyers.do"** uses the ACS survey data to prepare the potential buyers dataset;
  - **"analysis/0-Products-Prices.do"** uses Covered California Public Data to extract the set of plans and prices;

- **"analysis/4-First-stage.do"** estimates equation (8) and produces Figure 3(a), Online Appendix Table S1, and returns control functions for demand estimation;
- **"analysis/5-Demand.do"** estimates demand parameter via simulated maximum likelihood as detailed in Online Appendix S2.
- **"analysis/6-Cost.do"** estimates costs using Matlab as detailed in Appendix S2, produces Online Appendix S5;
- **"analysis/7-Postestimation.do"** computes willingness to pay, expected costs, elasticities, and produces Figure 5, and Online Appendix Figure S1;
- **"analysis/8-Demand-Inference.do"** computes standard errors for demand parameters, and produces Table 3 and Online Appendix Table S4;
- **"analysis/9-Cost-Inference.do"** computes standard errors for cost parameters, and produces Table 4, Figure 4, Figure 6, and Online Appendix Table S6;
- **"analysis/10-Counterfactuals.do"** computes equilibrium counterfactuals using Matlab. Produces Table 5, Table 6 and Figure 7;
- **"analysis/11-Summary.do"** produces all summary exhibits: Table 2, Figure 1, Figure 2, Figure 3(b);
- **"analysis/12-MoralHazard.do"** uses Matlab to estimate costs for various moral hazard parameters, produces Figure 8 and Table 7;
- **"analysis/13-AppendixDemand.do"** produces large tables with all demand parameters for online appendix, Tables S2 and S3.
- **~/Scripts/DO/compile surveys** contains Stata .do files obtained from IPUMS and MEPS to compile the publicly available surveys.
- **~/Scripts/M** contains all the Matlab scripts and functions invoked directly from Stata. These are used to estimate cost, simulate equilibrium counterfactuals, estimate cost with moral hazard, and simulate equilibrium counterfactuals with moral hazard.

## Instructions to Replicators

- Download entire folder, and execute `~/Scripts/DO/MASTER_DO_replication.do`
- Allow for up to 14 days of runtime using 22 cores and 256 G of RAM
- NOTE: **analysis/5-Demand.do** , **analysis/10-Counterfactuals.do**, and **analysis/12-MoralHazard.do** can be particularly time consuming, depending on resources.

---

---

## Data References

- [1] Agency for Healthcare Research and Quality (n.d.) *Medical Expenditure Panel Survey*. The Department of Health & Human Services. Obtained from <https://meps.ahrq.gov/mepsweb/>
- [2] Centers for Medicare & Medicaid Services (2013) *Market Rules Technical Summary*. U.S. Centers for Medicare & Medicaid Services. Obtained from <https://www.cms.gov/ccio/resources/files/downloads/market-rules-technical-summary-2-27-2013.pdf>
- [3] Centers for Medicare & Medicaid Services (n.d.) *Rate Review Data*. U.S. Centers for Medicare & Medicaid Services. Obtained from <https://www.cms.gov/marketplace/resources/data/rate-review-data>
- [4] Covered California (n.d.) *Data and Research*. The California Health Benefit Exchange. Obtained from <https://hbex.coveredca.com/data-research/>
- [5] Dartmouth Atlas Data (n.d.) *The Dartmouth Atlas of Healthcare*. The Trustees of Dartmouth College. Obtained from <https://data.dartmouthatlas.org/>
- [6] Internal Revenue Service (2015), *Instructions for Form 8962*. Department of the Treasury. Obtained from <https://www.irs.gov/pub/irs-prior/i8962--2015.pdf>
- [7] Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. (2015) *Integrated Public Use Microdata Series: Version 6.0* [dataset]. Minneapolis: University of Minnesota. Obtained from <http://doi.org/10.18128/D010.V6.0>
- [8] U.S. Census Bureau. (n.d.). *Relationship Files*. U.S. Department of Commerce. Obtained from <https://www.census.gov/geographies/reference-files/2010/geo/relationship-files.html>

## Appendix: Public Records Act Request

To whom it may concern,

For academic research purposes, under the Public Records Act (Govt. Code Section 6250 et seq.) I am requesting access to more detailed enrollment data from the initial open enrollment period of Covered California (from October 1, 2013- March 31, 2014) than what provided at the current Data & Research link: <http://hbex.coveredca.com/data-research/> .

To be as specific as possible, I would like the count of enrollees for every Rating Region, Insurance Carrier, Tier, Subsidy Status, and Age (at least in categories such as 20-25, 25-34, 35-44, 45-54, 55-64).

For example:

- *Region 1 , Tier: Bronze, Carrier: Health Net, Age: 20-24, Status: Subsidized, Enrollment: 2,300*
- *Region 1, Tier: Bronze, Carrier: Health Net, Age: 20-24, Status: Unsubsidized, Enrollment: 340*
- *Region 1, Tier: Bronze, Carrier: Health Net, Age: 25-34, Status: Subsidized, Enrollment: 3,000*
- *etc.*

This for **all** (region, age group, carrier, subsidy status, tier) **combinations**.

If possible, the exact plan (PPO/HMO/HSA) would ensure an even better quality of this data. I am affiliated with an educational institution, and this request is made for a scholarly purpose and not for a commercial use.

If you have any questions about handling this request, you may telephone me at (650)798-7705 or email at [ptebaldi@stanford.edu](mailto:ptebaldi@stanford.edu)

I request a **written response within 10 business day**. As mandated by law (§ 6253(c)).

Sincerely,

-----  
Pietro Tebaldi | Ph.D. candidate  
Department of Economics | Stanford University  
Landau Economics Bldg. | 579 Serra Mall  
Stanford, CA 94305-6072  
(650)798-7705 | [ptebaldi@stanford.edu](mailto:ptebaldi@stanford.edu)