

# **Supplementary material for “The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D”**

**by Liran Einav, Amy Finkelstein, and Paul Schrimpf**

This is a readme file for the supplementary material that accompanies “The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D” by Liran Einav, Amy Finkelstein, and Paul Schrimpf. Please contact any of us ([leinav@stanford.edu](mailto:leinav@stanford.edu); [afink@mit.edu](mailto:afink@mit.edu); [schrimpf@mail.ubc.ca](mailto:schrimpf@mail.ubc.ca)) with any questions.

The enclosed material includes program files that generate all the results reported in the paper. Note that the data used for the paper is confidential and cannot be made publicly available. They were obtained under DUA #22559 through NBER (PI: Amy Finkelstein). You must contact CMS (<http://www.resdac.org>) to apply for access to the data. If you are an NBER affiliate, contact Jean Roth at NBER: [jroth@nber.org](mailto:jroth@nber.org).

The program files are organized in three subfolders, the content of which is described in more detail below. In the “Descriptive Analysis” subfolder, we include two Stata do files that clean the data and produce the descriptive results reported in the paper. In the “Model Results” subfolder we include the program files and output files that underlie the model-based results. In the “Risk Scores” folder we include all the files that are used to generate the risk scores used in the baseline sample.

## **Descriptive Analysis subfolder**

### **qje\_DataPreparation\_final.do**

This file reads in the raw beneficiary, plan, and claim files for years 2007-2010 from CMS and ResDac and creates the baseline file that we use for analysis in our project. It goes through the following steps:

1) Cleaning Beneficiary files - it takes in raw beneficiary files, and limits the file to beneficiaries enrolled in Medicare Part D. It then cleans and labels various variables. We also generate variables for beneficiary birth month, join month, and dummies for whether a beneficiary switches plans during each year, and for whether he or she receives any cost-sharing.

2) Cleaning claims files - it takes in raw claims files with observations at the claims level. It collapses these files at the beneficiary level, creating variables for the dollar amount and number of claims at each phase of the standard Part D plan and for categories of drugs. We also generate monthly claim totals, average fill size, weekly spending variables, and the final phase that each beneficiary ends up in each year.

3) Cleaning plans files - Renames and cleans various variables within the plan data.

4) Merging 1), 2), and 3) above - we merge the beneficiary, claims, and plan files that have been cleaned above, and drop observations corresponding to claims unmatched to beneficiaries, claims not matched to plans, and plans without any claims. We drop any beneficiaries not in the strict 20% Medicare sample, and beneficiaries not in the US.

5) Appending each of the year-level datasets - We append the 2007-2010 year-level datasets generated by 4) above, then drop any duplicate beneficiaries randomly. We then drop beneficiaries who join for a reason other than turning 65, dual Medicaid eligibles, beneficiaries with listed 3rd party cost-sharing, non-stand alone Medicare Part D plans, beneficiaries who switch plans within the year, and beneficiaries who die within the year.

6) Creating baseline analysis variables - we take the dataset created by 5) and create variables for plan type, the yearly standard ICL, the yearly standard deductible, the difference between beneficiary spending and the ICL, deductible, and categorical spending amounts. We calculate beneficiary cost-sharing averages within each phase, and merge in risk scores. This is the baseline file, in which each observation is a beneficiary-plan-year for the years 2007-2010. We also generate a wide version of the baseline file for use in the cross-year substitution analysis.

We have two output datasets that result:

A) bene07080910\_baseline - long dataset in which observation is a beneficiary-plan-year for years 2007-2010

B) bene\_07080910\_wide - wide dataset in which each observation is a beneficiary-plan.

### **qje\_figures\_final.do**

This file reads in the two baseline data files created above and generates Figures II, III, IV, V, IX, A2, A3, A4, A5, and A9, and Tables I, II, III, IV, and A1. It begins by defining a series of programs that calculate the excess mass near the kink using the fitted polynomial table that is described in the paper, and the difference between fitted probabilities at the kink and the actual observed probability of claims. It also generates a version of the dataset that is used for cross-year analysis in the figures and tables, which reads in the wide dataset, and generates variables for the heterogeneity analysis and excess spending in January for the previous year.

## **Model Results subfolder**

“estimation” contains code for estimation.

“dataProcessing” contains code for preparing the data for estimation.

“counterfactuals” contains code for counterfactual simulations and reporting results

“deps” contains external dependencies.

To recreate the results:

1. Obtain access to the Medicare D data.

2. Prepare the data for estimation.

a) Compile Stata plugin for creating MySQL database by running “make” in the dataProcessing directory.

b) In Stata: “do createSQLite”

3. Run the estimation.

a) Compile the estimation program: run “make smm” in the estimation directory.

b) Execute “./smm path-to-data.sl3 baseline3.cfg” (modify path-to-data.sl3 and baseline3.cfg with the name of the data and configuration file you want to use); this may take a while (likely multiple days).

4. Run the counterfactuals and create the results.

a) Compile the simulation program: run “make cf.so” in the counterfactuals directory.

b) Modify “run.R” as needed and execute it in R

c) The results will be compiled into a pdf based on tex/slides/slides\_PartD\_paulFigures.tex

## **Risk Scores subfolder**

These files must be run in the order they are listed below:

### **Contents.sas, DataMacros.sas, UtilityMacros.sas**

Various macros for reading in data that are used to facilitate programming in the SAS files below. These files are not independently run.

### **BeneDataPrep.sas**

This file reads in raw beneficiary data from Medicare from 2006-2010 and cleans the demographic and enrollment information, to input into SAS risk score programs.

### **DiagnosesDataPrep.sas**

This file reads in raw diagnoses from Medicare from 2006-2010 and cleans the diagnosis names and ICD-9 code information, to input into SAS risk score programs.

### **RunRxRiskScores20xx.sas**

These files create RxHCC risk scores in SAS files for each respective year, based on CMS programs and formulas that are in subfolder RxHCC\_2012software. Please see the CMS documentation in this folder for more detail on the risk score software.

### **RunMARiskScores20xx.sas**

These files create MA HCC risk scores, based on CMS programs and formulas that are in subfolder CMSHCC\_2012software\_MA\_Cost. Please see the CMS documentation in this folder for more detail on the risk score software.

### **Specialists.sas & Specialists\_Collapse.do**

These files create a crosswalk between the beneficiaries in baseline sample and the specialties of the physicians that they are seeing, for use in the Medicare Parts A and B utilization file.

### **CodingMedicareUtilization.sas**

This file reads in Medicare Part A and B data, and data on specialists that our beneficiaries see, to create variables on the utilization of Parts A and B and the death dates of our beneficiaries who die during 2006-2010, to merge into the baseline data.

### **cleaning\_scores.do**

This .do file reads in the risk score and utilization files generated above for each year, and then appends the yearly files to create 4 output files:

- A) MA\_riskScores\_baseline – HCC risk scores for our baseline beneficiaries in 2006-2010
- B) RiskScores\_baseline – RxHCC risk scores for our baseline beneficiaries in 2006-2010
- C) utilization\_tomerge – Utilization of Medicare Parts A and B for our baseline beneficiaries in 2006-2010
- D) death\_dates\_full – Death dates of beneficiaries that died in our sample.