Replication Guide for "Optimal Long-term Health Insurance Contracts: Characterization, Computation, and Welfare Effects

March 28, 2023

This document describes how to replicate the empirical analysis in the paper. We provide these descriptions in three sections: (i) data availability (ii) data construction (iii) detailed replication instructions. Please contact either Soheil Ghili (soheil.ghili@yale.edu) or Ben Handel (handel@berkeley.edu) with questions about the paper or the replication materials.

1 Data Availability

For our analysis we construct key inputs for our model (described below in "Data Construction") based on the Utah All-Payer Claims Database. This database is administered by the Office of Health Care Statistics for the state of Utah and is only able to be accessed by researchers who go through the data approval process with the Office and have an executed data-sharing agreement. The Office of Health Care Statistics outlines the clear process for accessing these data on their website:

https://stats.health.utah.gov/about-the-data/data-series/

In order to access these data please follow the process listed on this website. In addition to the data provided by the state of Utah, we leveraged derived variables related to consumer health status that used the Johns Hopkins ACG Clinical Software program. This program can be found here:

https://www.hopkinsacg.org/

This program inputs claims and demographics from the Utah data and derives new variables related to overall health status and predicted health status with its proprietary algorithm. To use this software in conjunction with the Utah data, or otherwise, please contact the software administrators at the above website and execute a software use agreement.

2 Software & Runtime

Our replication package, as described in detail below, can be run entirely in Matlab using the inputs provided. Approximate runtime is one hour.

3 Data Construction

Our primary analysis uses inputs from the Utah all-payer claims data merged with the derived ACG health prediction variables. As described in the text, there are three basic sample restrictions in our main analysis:

- Restrict to consumers present in the data for at least 8 months in 2013, 2014 and 2015
- Restrict to men
- Restrict to those with ages between 25-64 for all years 2013-2015

For each individual, the Johns Hopkins algorithm produces a predicted risk-score for future spending based on an individual's claims, demographics, and spending from a calendar year. Thus, each individual has one predicted risk score based on each year of data (2013 data predicting 2014, 2014 predicting 2015, and 2015 predicting 2016). We separate the continuous predicted spending variable in 7 equally-sized septiles that partitions the population into groups with different levels of predicted health risks. We provide two data matrices based on these risk-scores and transitions as key inputs into our analysis, described in the next several paragraphs.

The main data file used in the analysis is stored in a file titled "menData2.mat" in the main folder. It includes two main variables. One is titled HCC and stores the expected annual healthcare expenses for each of the seven possible health states. Remember that we have a total of $49 = 7 \times 7$ possibilities here, given that we allow for the health status of the patient this year and in the past year to affect future transitions. Nevertheless, as far as the current year's expected healthcare expenses are concerned, only the current health status matters. As a result, the 1×49 HCC vector is basically a 7-times repetition of a 1×7 vector.

The second variable stored in this data file is the transition matrices. Recall that our model is a second order Markov process and we have separate transition processes for different 5-year age chohorts. Also recall that there are 8 such 5-year age chohorts in our setting (between 25 and 64 years old). As a result, one would naturally expect a $49 \times 7 \times 8$ tensor consisting of 8 transition matrices each of size 49×7. That said, the data has a $49 \times 49 \times 8$. This is simply because we re-parameterize and capture the 2-nd order Markov process with 7 states as a 1st order one with 49 states.

The first three tables in the main paper describe the information contained in these two data matrices.

In addition to these two key matrices (health costs per health status group and 2nd order health status transitions by age cohort) we bring in other elements to our simulations from outside sources as cited in the main text. These include (i) the discount factor (ii) the initial age-25 health status distribution (iii) the CARA risk coefficient (iv) and lifetime income profiles. All of these data elements are simple and are, hence, directly incorporated in our main code file titled "Main.m" which will be discussed at length in the next section that describes how to replicate our analyses of the shape of the optimal dynamic contract and consequent welfare results.

4 Detailed Replication Instructions

Tables 1 and 2: These are direct summaries from the data which is not available to the public reader.

Table 3 top panel: Go to path "MS_files/source/persistence". Run the file "TransitionTables.m". Results will be saved in path "MS_files/output/markov2_transitions" in multiple .csv files. The table reported is a manually shortened version of all of those results.

Table 3 bottom panel: Go to path "MS_files/source/persistence". Run the file "PersistenceTables.m". Results will be saved in path "MS_files/output/persistence". Results are saved in multiple files. See in particular, "tab30_markov2_grand.csv". The reported table is an abridged version of this csv file. It only contains rows (1,1), (4,4), and (7,7) (written 101, 404, and 707 respectively) from the csv. Also note that the csv has the total expenditures while the table in the paper has converted them to average expenditures. To see this, note that the first columns match. To match the second column, the result in the csv should be multiplied by $\frac{1-\delta}{1-\delta^5}$. Same holds for the third column. For the fourth and last column, the number to multiply is $\frac{1-\delta}{1-\delta^{34}}$. In all of these, δ is the discount factor and set to 0.975. The output file "tab30_markov2_grand.csv" includes a version of the table to which this averaging operation has been applied.

Figure 2: To replicate, navigate to " "MS_files/source/income_streams" and run the script "PlotIncomeStreams.m". Running the script generates Figure 2, which is saved under "MS_files/output/income_plot/income_profiles.png".

Table 4: The top panel presents first-year equilibrium contract features for flat net income. To replicate, go to path "MS_files/source/contract_analysis". Run the file "ContractTerms.m". (Note that this file take a relatively long time to fully run.) Results will be saved to "MS_files/output/contract_tables/FN2_markov2-first_year.csv".

The bottom panel is the produces the same result but for the downscaled manager group. This table gets produced by running the same file (thus, no need to rerun it). The results are saved in the same directory as the results file stated in the preceding paragraph under the filename "DM_markov2-first_year.csv"

Table 5: This table presents second year premiums for some health status histories under the flat net income. This table is generated using the same procedure as table 4. Thus, there is no need to run additional code. The results are saved under "FN2_markov2-premiums.csv". Note that the table in the paper is an abridged version of the csv file, including only rows 1, 4, and 7.

Table 6: This table provides welfare results for the flat-net and downscaled-manager income paths as a function of initial health history. To generate this table, navigate to "MS_files/source/init_state". Then, run the script "InitialHealthState_FlatnetM2_Welfare.m". The script saves Panel A of the table to "output/init_health/welfare_m2_small-FlatNetM2.csv" and Panel B of the table to "output/init_health/welfare_m2_small-DownsMngr.csv".

Table 7: This table provides welfare results from the perspective of an "unborn" consumer who has not yet observed his health status. To replicate, go to path "MS_files/source/contract_analysis". Run the file "WelfareTables.m". Results for columns (1), (2), (3), (4), (6), (7), and (8) will be saved at "MS_files/output/welfare_tables/markov2_empirical.csv" (note that the script takes a few minutes to run). To obtain the "Req. subsidy" reported in column (10), run the script "MS_files/source/init_state/InitHealthState_Markov2_sim.m". This script saves results in the directory "MS_files/output/init_health". The output files take the form "markov2_sim_small-*.csv", where the * stands for the name of the various income profiles. These files contain the values of Table 7's column (10) in their entries under the "Ex ante, (empirical)" row and the "Subsidy" column. To obtain column (5) of Table 7, navigate to "MS_files/source/init_state" and run the script "InitHealthState_Markov2_sim_Marginal.m". This script saves results in the directory "MS_files/output/init_health". The output files take the form "MU_markov2_sim_small-*.csv", where the * stands for the name of the various income profiles. These files contain the values of Table 7's column (9) in their entries under the "Ex ante, (empirical)" row and the "BESubsidyEmp" column. The values of column (9) are straightforwardly computed with the results in columns (5), (3), and (2) using the formula displayed in the header of column (9).

Table 8: To replicate, navigate to the directory "MS_files/source/aca" and run "ACA.m". Running this script generates the output file "MS_files/output/ACA/baseline/markov2_emp.csv", which contains the results displayed in columns (1), (2), (3), and (5) of Table 8. The results in column (7) may be straightforwardly computed using results in the "markov2_emp.csv" file in addition to the formula in the header of column (7). To obtain column (4) of Table 7, navigate to "MS_files/source/init_state" and run the script "InitHealthState_Markov2_ACA_Marginal_sim.m". This script saves results in the directory "MS_files/output/init_health". The output files take the form "MU_markov2_ACA_big-*.csv", where the * stands for the name of the various income profiles. These files contain the values of Table 8's column (4) in their entries under the "Ex ante, (empirical)" row and the "BESubsidyCE" column. The values of Table 8's column (6) are straightforwardly computed with the results in columns (4) and (2) as well as the CE_{SPOT} results in the "MS_files/output/ACA/baseline/markov2_emp.csv" file using the formula displayed in the header of column (6).

Table 9: To replicate, navigate to the directory "MS_files/source/switching_cost_Utah" and run "SwitchingCost_empirical_init.m". Running this script generates the output file "MS_files/output/switch_tables/switchcosts-emp_init-markov2.csv", which contains the results displayed in Table 9.

Table 10: To replicate, go to path "MS_files/source/contract_analysis". Run the script "WelfareTables.m". The script saves results to "MS_files/output/welfare_tables/markov2_empirical_lowRA.csv" (note that the script takes a few minutes to run).

Figures 3, 4, and 5. To produce these figures, navigate to "MS_Files/source/myopia". Run two MATLAB files titled "ProduceMyopiaTables.m" and "ProduceLapsationTables.m". The order in which you run the files does not matter. The first MATLAB file produces a csv file titled "'results_Myopia.csv" while the second file generates the files "*LapsationResults_Myopia.csv", where "*" is the name of an income profile; these files are saved in the "MS_files/output/myopia" directory. Manually copy these two csv files into the folder "MS_files/source/Revision Graphs". Then run the files "Myopia.R" and "LapsationRates.R" on a R.4.1.2 or a more recent version. Among the images produced by running these files, you may find "myopiaFrontloads.pdf", "lapsationsdownsManager.pdf", "meanLapsations.pdf", and "myopia.pdf". These are, respectively, Figure 3, left panel of Figure 4, right panel of Figure 4, and Figure 5.

5 Data References

Utah All Payers Claim Database, 2013-2015. Utah Health Data Committee/Office of Health Care Statistics. Utah Department of Health. https://healthcarestats.utah.gov/about-the-data/apcd/

Anonymous Firm. 2004-2009. "Health care claims, choice and demographics data for large employer." Unpublished data.