“The Early Impact of the Affordable Care Act State-By-State”

Amanda Kowalski

README for Empirical Analysis Replication Files

This file serves as a guide to the empirical analysis conducted in the paper and executed by the replication files. It is organized into four sections: **Directory & Program Structure**, **SNL Data Download**, **Raw Data Processing**, **Graphical Analysis**, and **Welfare Analysis & Regressions**. The first section describes the file structure of all programs and raw inputs. The second section provides explicit instructions that may be used to download the raw data from SNL, conditional upon having a paid subscription agreement with SNL. Our data use agreement with SNL does not allow us to release the raw data. Accordingly, **we do not provide the raw SNL data files** with these replications files; however, we do include data files which contain aggregate statistics derived from the underlying SNL data. The third section describes the programs that clean and process the raw data. The fourth section describes the programs that construct all graphs that appear in the paper, appendix, and presentations. The fifth and final section describes the programs that conduct the welfare analysis and construct the tables that appear in the paper and appendix.

1. Directory & Program Structure

The directory structure is as follows:

- All .do files (including "master.do") are saved in the main directory. The main directory also serves as the working directory, referred as "wd" here. All file paths are relative file paths to "wd".
- Raw data are saved in "data/raw"; Intermediate data and processed data are saved in "data/intermediate" and "data/processed", respectively
    - Because our data use agreement does not permit us to release the raw SNL data, we include only intermediate and processed data files with which to run the analysis
    - Consequently, the files described in Section 3 cannot be run unless the raw data are obtained from SNL by means of a paid subscription agreement with SNL
- All other directories are created by "master.do". They include:
    - "data/intermediate", which contains intermediate data sets created in Stata
    - "data/processed", which contains processed data sets created in Stata
    - "logs", which contains all log files
    - "graphs", which contains all graphical output
        - Contains the subdirectories "graphs/paper", "graphs/appendix", and "graphs/presentation"
    - "output", which contains all final output
        - Contains the subdirectories "output/regs" and "output/tables"
        - "output/tables" contains output used in the construction of tables that appear in the paper and online appendix
        - "output/regs" contains regression output, which is divided into "output/regs/regs_joint" and "output/regs/regs_univar", accordingly

The file "master.do" is located in the working directory, **which is to be specified by the user** (****In order to run the programs, this must be changed**).** The file "graphs_presentation.do" is the only file that may not be run in batch mode. This must be run using the Stata GUI. The other two graphs files ("graphs_paper.do" and "graphs_appendix.do") may be run in batch mode.

2. SNL Data Download

As stated above, our data use agreement with SNL does not permit us to release the raw data files. We do, however, provide intermediate data files which contain aggregate statistics from the underlying SNL data, and these are adequate to run our programs from an intermediary stage. For those with the appropriate paid subscription agreement with SNL, the following instructions will allow you to download the raw data files from SNL.

Register on the SNL website to download the SNL add-in for Excel. Then follow these steps:
[Note: steps 1-3 and 7 are the same for all three databases (quarterly data, annual data, and schedule T data). Steps 4-6 are specific to each file.]

1. Select “SNL Financial” tab in Excel toolbar. Select “Data Wizard” and log in.

2. Under "Search Criteria”, choose the following settings:
   - File: US insurance statutory financials
   - Entity Type: Health
   - Operating status: Operating
   - Reporting level: Individual Cos
   - Entity type: Companies

3. Under “Report Builder”:
   - Select the group “Corporate” and the subgroup “Address and geographic”. Add “State domiciled” to the export list.
   - Again under “Corporate”, select the subgroup "Group/parent data". Add “SNL Group name" (name of company) to the export list.

4. *[Note: the following is for quarterly data]*
   - Select the group “State level data” and the subgroup “State page totals”. In the drop-down menu (located above the time-period tab), select “Quarter”. Then select the quarters Q2 2014 through Q1 2008.
   - In the "Lines of Business" drop-down menu, select “AR: Comprsv (Hosp, Med) Ind" (Individual health insurance).
   - In the "Fields" tab, select “Total Health Enrollment", “Member months", "Direct premiums written", and "Provision of Health Care Services Paid". Add these to the export list.
   - On the next page, in the application tab of the *settings* menu, select “list periods down a column”.
   - Select “ok” and “export SNL table”.
   - Save as "quarterly_dat_nd.csv".

5. *[Note: the following step is for annual data – indications are for year 2008 but proceed in the same way for years 2009-2013]*
   - Select the group “state level data” and the subgroup “Stage page (annual)”.
   - In Time period tab, choose 2008.
   - In the “Line of business” dropdown, select "AR: Comprsv (Hosp, Med) Ind" .
   - Select all states (“AK” to “WY”).
   - In the field column, choose “Total Health Enrollment", “Member Months”, “Direct Premiums Written” and “Provision of Health Care costs paid”.

- On the next page, in the “application tab” of the *settings* menu, select "List entities down column". Select “ok” and “export SNL table”.
- Save file as “2008_nd.csv”.

6. *[Note: the following step is for Schedule T data]*
   - Select the group “state level data” and the subgroup “Schedule T premiums”.
   - In the time period tab, choose “Quarter”. Then select the quarters Q2 2014 through Q1 2008.
   - Under “Line of Business”, select “AR: health (A&H Sch T)”.
   - Under “Geography”, select all states (“AK” to “WY”).
   - Under “Fields”, select “Direct Business less Emp Ben contributions”.
   - On the next page, in the “application tab” of the *settings* menu, select "List entities down column". Select “ok” and “export SNL table”.
   - Save file as “scheduleT.csv

3. Raw Data Processing

The raw data is processed in the files “clean_data.do”, “impute_data.do”, “state_alloc.do”, “data_collapse.do”, and “misc.do”. The main data used in the analysis have been downloaded from SNL. These data are proprietary and may not be shared. Though these data have not been included in the replication files, we have included the relevant programs. Their functions are as follows:

- “clean_data.do” :
  - Imports raw data from native format and saves as .dta. This includes raw data from other sources, including the Census, ASPE reports, and the Kaiser Family Foundation.
- “impute_data.do” :
  - Performs imputation routine to correct apparent data errors in the raw data from SNL. Details of the imputation procedure are provided in the appendix to the paper.
- “state_alloc.do” :
  - Allocates data from insurers to the state level, relying on annual data and Schedule T data when necessary. Details of the allocation procedure are provided in the appendix to the paper.
- “data_collapse.do” :
  - Collapses the data to the state level and to the state group level. Merge data on various state policies that may have affected the state level impacts of the ACA.
- “misc.do” :

- Computes various statistics and figures that are reported in the paper. Assesses the magnitude of the imputation procedure (*i.e.* how many/how important were the observations that were imputed).

4. Graphical Analysis

These files construct the graphs that appear in the paper, online appendix, and in the presentation of the paper.

- "graphs_paper.do" :
  - Constructs graphs of enrollment, premiums, and average costs of state groups and of major states.
- "graphs_appendix.do" :
  - Constructs graphs by state that compare imputed data to raw data for all variables (enrollment, coverage, premiums, and average costs).
- "graphs_presentation.do" :
  - Constructs graphs of enrollment, premiums, and average costs of state groups and of states. Formatting differences from "graphs_paper.do" to display better on PowerPoint.

5. Welfare Analysis & Regressions

These files implement the welfare analysis and assess the impacts of state policies on changes in welfare. Their purposes are as follows: apply the model developed by Hackmann, Kolstad, and Kowalski ("HKK") in order to compute the change in welfare

- "welfare_results.do" :
  - Implements the model described in the paper to compute enrollment, premiums, and average costs had the ACA not been implemented and all states followed their seasonally adjusted trends. Calculates the change in welfare as a result of the ACA at different levels of the penalty.
  - Generates "output/tables/welfareresults_k.csv" for k = 1,2,3,4. This output is used in the construction of all tables other than the regression tables, including Table 1, Table 2, and Table A1 (and their analogues in the online appendix).
- "regs_joint.do"
  - Runs welfare regressions, controlling for all state policies simultaneously. Constructs bootstrapped confidence intervals for all coefficients.
- "regs_univar.do"
  - Runs welfare regressions, examining the effect of one state policy at a time. Constructs bootstrapped confidence intervals for all coefficients.

- "generate_tables.do"
    - Compiles results from "regs_joint.do" and "regs_univar.do" into a format that may easily be used to construct the regression tables, including Table 3 (and its analogues in the online appendix).
    - Generates "output/tables/reg_results_joint.csv" and "output/tables/reg_results_univar.csv"

Each variation of the analysis (1, 2, 3, and 4) refers to a distinct iteration. The first (k=1) is the base model, relying on the imputed SNL data and a post-reform period extending over both Q1 and Q2 of 2014. The second (k=2) is similar to the base model, differing only in that it relies on the *raw*, rather than imputed, SNL data. The third (k=3) again relies on the imputed data, but this variation uses only Q1 2014 for the post-reform period. Finally, the fourth (k=4) uses the imputed data but uses only Q2 2014 for the post-reform period. Thus, for example, the estimated regression coefficient for direct enforcement in the univariate regressions using *k=1* and *pi=1000* is in the column "reg_1_1000".