Exercise 1 Solutions

Ian McCarthy, Econ 771

Fall 2020

Overview

This document lays out my approach to addressing the questions in Exercise 1. There are lots of ways to answer these problems and build the data, especially in `R`! My way of answering the questions is just one of many.

I also have a particular workflow that you may or may not choose to adopt. That workflow is to do all of my analysis seperately from my markdown document. I really don't like seeing huge markdown documents constantly going in and out of data work and discussion. It's also not feasible to work with really large data within the markdown document. So, I do all of the analysis, remove any large objects (including the data), keep only the relevant objects for the final markdown (figures, tables, specific numbers or statistics), and then save the workspace. Then I load the workspace into my markdown document.

Data

Before we get into the specific questions for this assignment, let's briefly go through the data. My code files for the IPPS and POS data are included in the classroom repo under the filenames *1_ipps.R* and *2_pos.R*.

IPPS Data

The IPPS data are pretty straightforward except for some slight changes in variable names over time. For example, "Provider Type" is missing in years 1986 through 1983; it is named "provtype" in years 1994 through 2000; and it is renamed "ptype" from 2001 onward.

POS Data

The POS data may be a little troublesome to download because there are lots of different versions of the same datasets. From the NBER POS website, I extracted the "csvYYYY" files from 1984 through 1990. Starting in 1991, there are two links for the csv files – a "csv" and a "PROV". The "PROV" link is active but the datasets are empty in some of the earlier years, so I used the "csv" links in all cases. These links are to a crosswalk version of the data, in which the NBER tried to assign a consistent name to each variable in the POS data. The original, raw POS files had variable names like, "prov0028", but they changed to mnemonic style names in 2011. The crosswalk files are intended to make it easier to merge historic data with data from 2011 onward. These versions of the data are what I have made available in the AWS AMI.

The POS data are supposed to be updated regularly, meaning that you should only need the most recent version of the data; however, that doesn't appear to be the case as some provider numbers only show up in the earlier years. It appears that once a hospital has been closed for a sufficiently long time, it no longer appears in the POS. Still, since the values are intended to be replaced every year for a given POS, there are **a lot** of duplicates. In my code file, I read in all of the data, recode the relevant variables, and then drop the duplicates.

Final dataset

Once you have the ipps and POS data ready, you just have to merge the two. The easiest way to do this in our case is to merge on provider number and year. Here's my code to merge the three datasets:

```
full.data <- final.pos.data %>%
  left_join(final.ipps.data, by=c("provider", "year"))
```

Question and Answers

Question 1.

Provide and discuss a table of simple summary statistics showing the mean, standard deviation, min, and max of the case mix index and disproportionate share across years. Show these statistics overall and separately by hospital ownership type.

Table 1 presents the mean, standard deviation, min, and max of the relevant variables. Note that

disproportionate share information isn't available in the data until 1994, so the summary stats are missing until that year. We see from the table that the mean case mix index is increasing over time, which means that people are getting sicker. We also see that the mean DSH has increased over time, from about 23% to 30% in 2018.

Question 2.

Create a figure showing the average hospital disproportionate shares from 1986 to 2018. Show this trend separately by hospital ownership type (private not for profit and private for profit).

A table is not the best way to present summary statistics over several years like this. So, let's do that in a figure instead. Figures 1 and 2 present the average Case Mix Index and disproportionate shares, respectively, for each year. Each figure presents two lines, one showing the mean among for profit hosptials and the other showing the mean for private not-for-profit hospitals (labeled simply as "Non Profit" in the figures).

Question 3

Using a simple DD identification strategy, estimate the effect of Medicaid expansion on the hospital disproportionate share percentage. Do these effects differ by profit status (i.e., re-estimate with a DDD using profit status)? Focus your analysis on the years from 2005 through 2018.

The disproportionate share program is actually two programs - a Medicaid and a Medicare program. The Medicare DSH program, which is what is relevant for the IPPS data, calculates a hospital's disproportionate share percentage based on the share of Medicare patients that are eligible for SSI as well as the number of Medicaid patients. Since not all states expanded Medicaid in 2014, we can use the ACA Medicaid expansion as a treatment and estimate the effect of Medicaid expansion on hospital DSH percentage (which should be positive, since more people would be on Medicaid).

Before thinking of any type of regression, it's always a good idea just to summarize the key parts of your data. In this case, we want to know how DSH changed from before and after 2014, separately by states that expanded and didn't expand Medicaid.

Table 2 presents the 2x2 DD chart with pre/post, treatment/control averages. From these averages, our DD estimate for the effect of Medicaid expansion on DSH would be 0.00687.

We can estimate this effect in a regression setting based on the specification in Equation (1):

$$DSH_{ht} = \alpha Expand_s + \beta 1(t \geq 2014) + \delta Expand_s \times 1(t \geq 2014) + \varepsilon_{ht}, \qquad (1)$$

where DSH_{ht} is the Disproportionate Share percent for hospital h at time t, $Expand_s$ denotes an indicator set to 1 if the hospital is part of a Medicaid expansion state and 0 otherwise, and $1(t \geq 2014)$ denotes an indicator for the treatment period (after the DRG change in 1988.

Since we have multiple years of data, a more general specification is presented in equation (2):

$$DSH_{ht} = \delta Profit_h \times (t > 1988) + \lambda_h + \gamma_t + \varepsilon_{ij}, \qquad (2)$$

where λ_h and γ_t denote hospital and year fixed effects, respectively. Hopefully it is clear that Equation (2) subsumes Equation (1) as a special case.

The estimates for δ from each specification are presented in Table 3. These estimates confirm our original estimates in the 2x2 DD table, where we find a relatively large increase in DSH among hospitals in expansion states

Question 4

Estimate an "event study" version of the DD model from part 3. Focus only on states that expanded in 2014 that those that have never expanded.

Event study estimates for 2014 expansion are presented in Figure 3.

Question 5

Estimate another event study where you allow for differential treatment timing and incorporate all states.

Event study estimates for all expansion states (time varying treatment) are presented in Figure 4. Here, time of 0 denotes the start of treatment, so *t-1* is the baseline period for which every other period is estimated. That's why the coefficient for the *t-1* period is normalized to 0. Note that the full data run from 2005 to 2018, so DSH in period *t-12* is only observed for states that expanded in 2016. Since very few states expanded in 2017 or 2018, I've lumped all of those states into the *t-12* period.

Tables and Figures

Table 1: Summary Statistics

	Case Mix Index				Disproportionate Share			
Year	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
1986	1.12	0.14	0.59	2.23				
1987	1.15	0.15	0.72	2.15				
1988	1.18	0.17	0.61	2.16				
1989	1.20	0.18	0.63	2.28				
1990	1.20	0.19	0.56	3.21				
1991	1.22	0.21	0.50	2.75				
1992	1.23	0.22	0.48	2.61				
1993	1.20	0.24	0.47	4.42				
1994	1.22	0.24	0.44	3.72	0.23	0.18	0	1.60
1995	1.22	0.25	0.41	2.87	0.24	0.19	0	1.66
1996	1.24	0.25	0.31	3.10	0.24	0.19	0	1.64
1997	1.25	0.26	0.41	3.67	0.25	0.19	0	1.28
1998	1.27	0.40	0.40	15.93	0.24	0.19	0	1.33
1999	1.27	0.35	0.42	16.34	0.24	0.18	0	1.30
2000	1.26	0.28	0.42	3.71	0.24	0.18	0	1.28
2001	1.25	0.28	0.42	3.83	0.24	0.18	0	1.52
2002	1.25	0.29	0.29	3.80	0.23	0.18	0	1.29
2003	1.26	0.27	0.39	2.99	0.24	0.18	0	1.34
2004	1.28	0.28	0.40	4.02	0.25	0.18	0	1.32
2005	1.30	0.29	0.36	3.53	0.26	0.18	0	1.24
2006	1.33	0.30	0.36	3.07	0.26	0.18	0	1.23
2007	1.36	0.31	0.41	7.52	0.27	0.18	0	1.24
2008	1.36	0.31	0.45	4.13	0.27	0.18	0	1.20
2009	1.36	0.31	0.58	3.74	0.27	0.18	0	1.22
2010	1.41	0.34	0.50	3.74	0.27	0.18	0	1.22
2011	1.43	0.33	0.57	3.75	0.28	0.18	0	1.24
2012	1.45	0.33	0.60	4.28	0.28	0.18	0	1.24
2013	1.46	0.33	0.59	3.76	0.28	0.18	0	1.19
2014	1.48	0.35	0.56	3.80	0.28	0.18	0	1.23
2015	1.51	0.35	0.65	3.97	0.28	0.18	0	1.22
2016	1.53	0.36	0.59	4.36	0.29	0.18	0	1.49
2017	1.54	0.37	0.61	3.97	0.30	0.18	0	1.23
2018	1.58	0.37	0.65	4.88	0.30	0.18	0	1.19

Table 2: DD Table for Medicaid Expansion

Group	Pre	Post
Non-expansion	0.28	0.27
Expansion	0.29	0.29

Table 3: Difference-in-Differences Estimates

	Disproportionate Share	
	(1)	(2)
After 2014	0.001	
	(0.003)	
Expansion	0.014***	
	(0.002)	
Treatment Effect	0.040***	0.035***
	(0.004)	(0.001)
Constant	0.268***	
	(0.002)	
Model	DD	DD, FE
Observations	42,698	42,698

Note:	*p<0.1; **p<0.05; ***p<0.01	
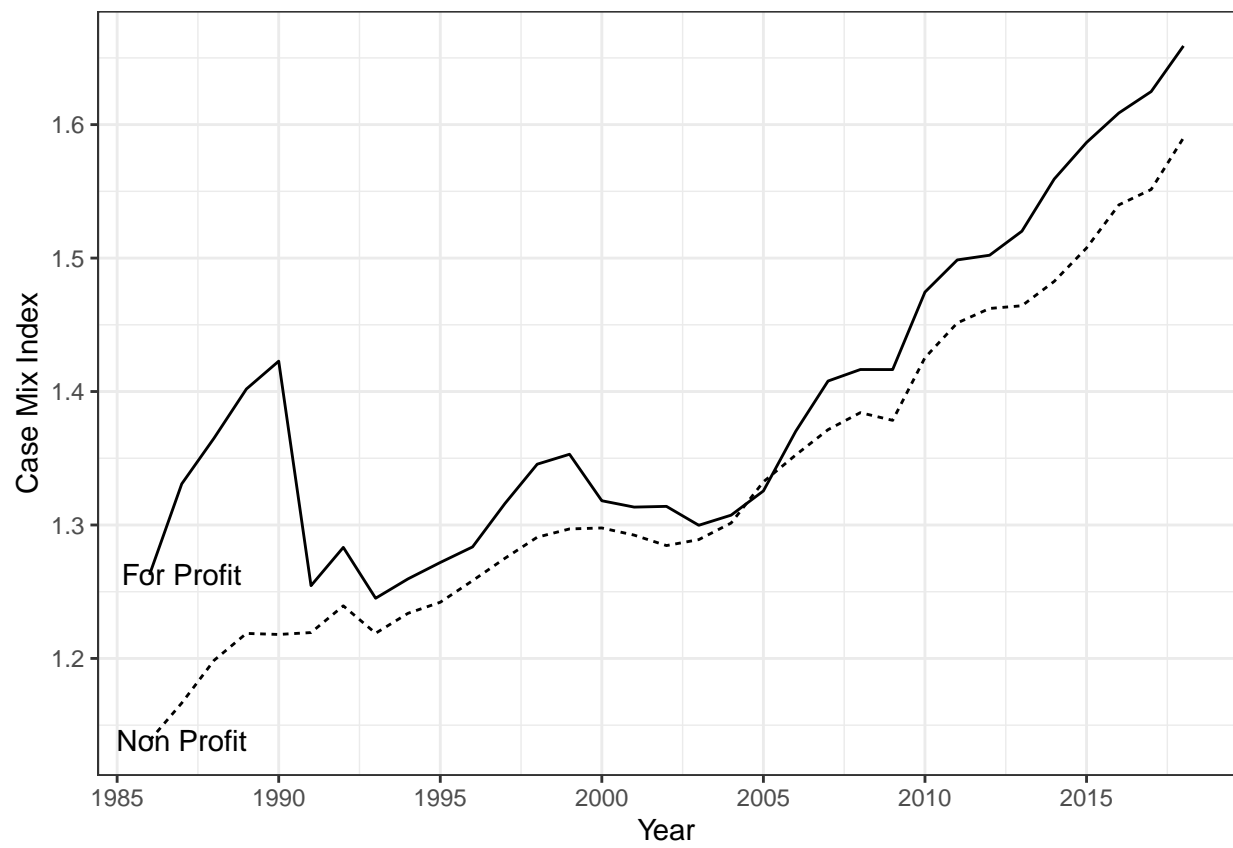
Figure 1: Mean Case Mix Index by Year and Profit Status
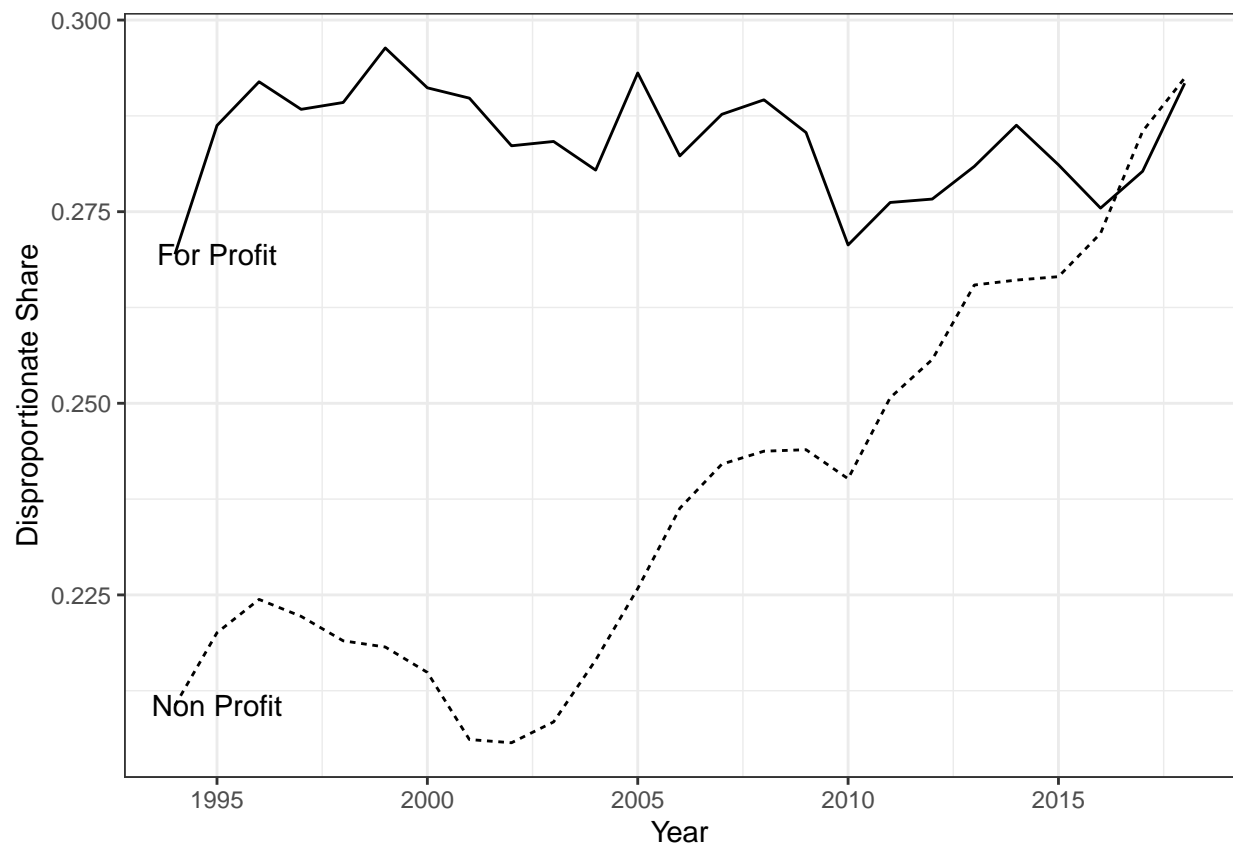
Figure 2: Mean Disproportionate Share by Year and Profit Status
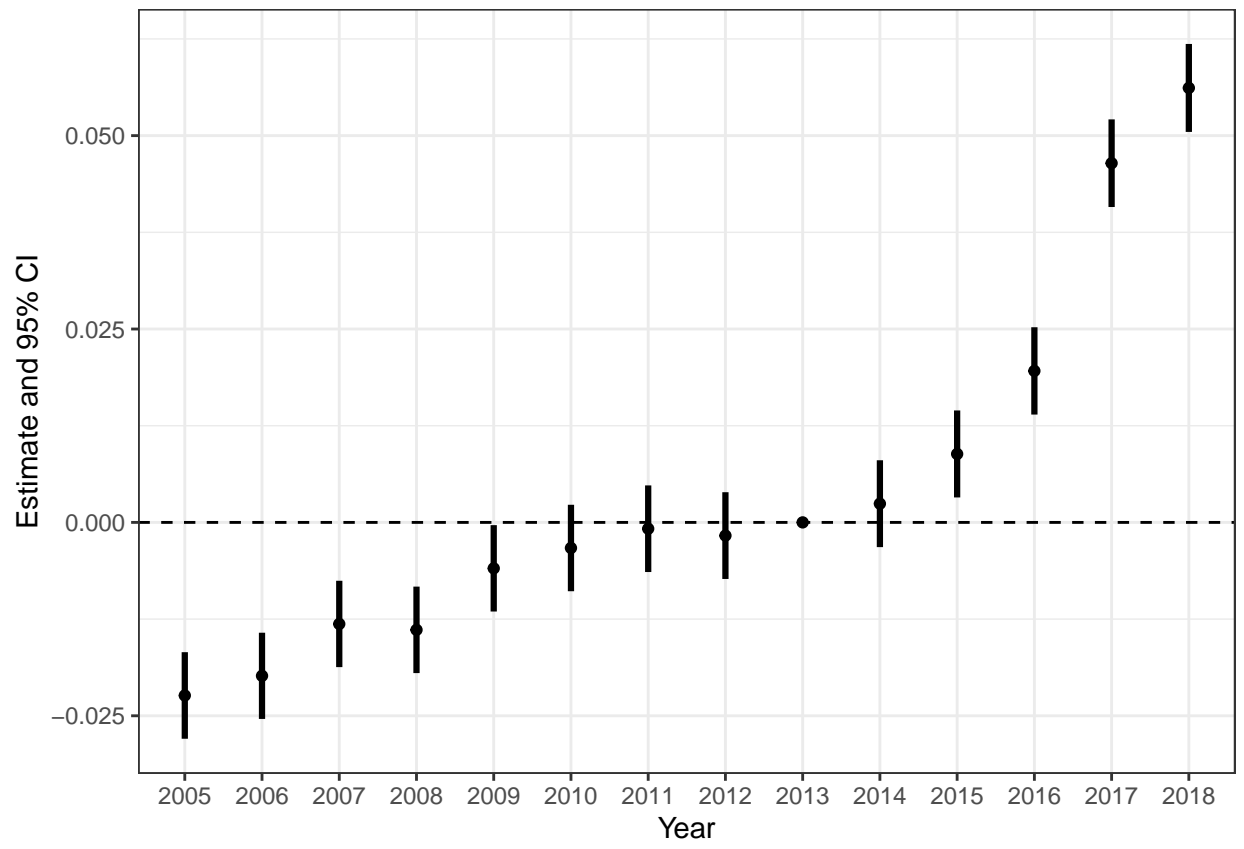
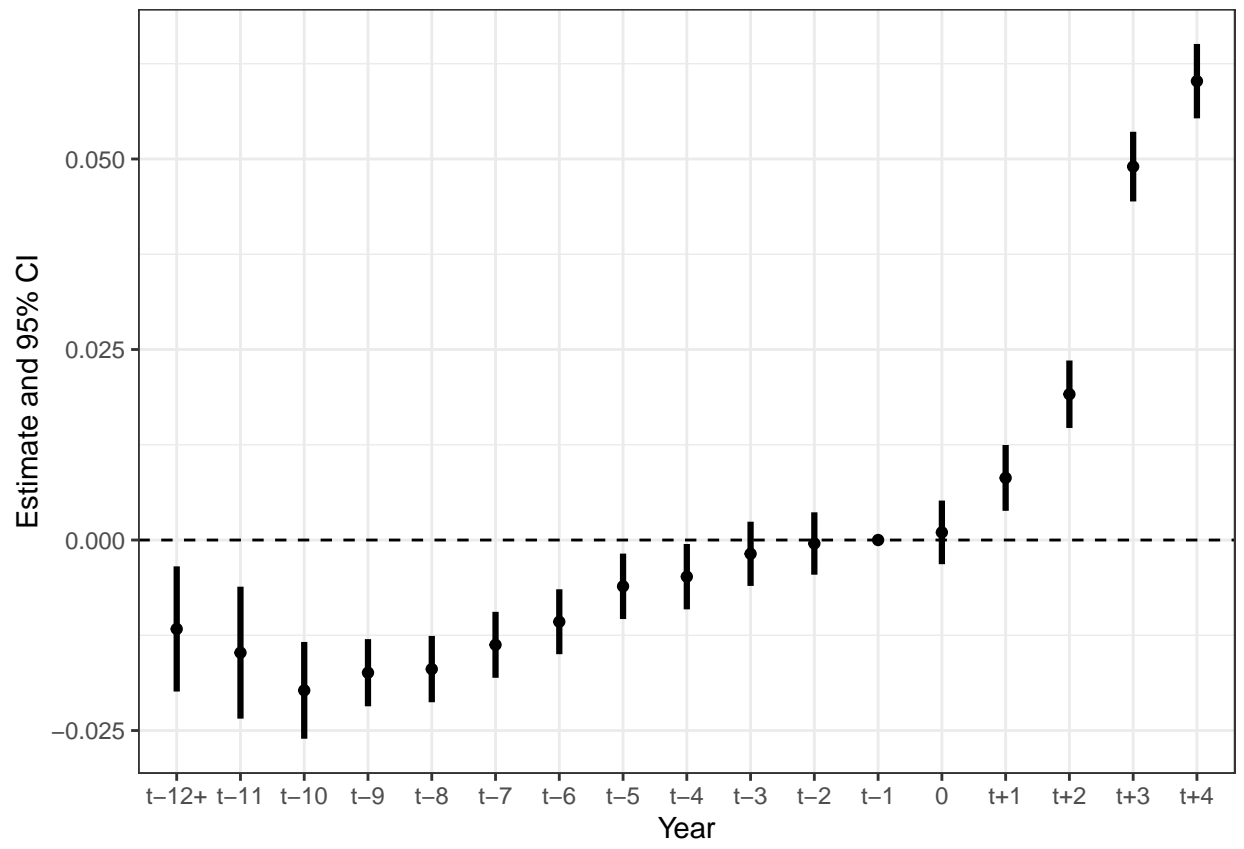Figure 3: Event Study for 2014 Treatment Group

Figure 4: Event Study for All Expansion States

References