Exercise 3 Solutions

Ian McCarthy, Econ 771

Fall 2020

Overview

This document lays out my approach to addressing the questions in Exercise 3. 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

The data in this assignment are straightforward since we're working directly off of Ericson (2014). All of the data are available on the journal website for the paper, here.

Still, it might take some time to understand the data and variable construction. The data are organized into two datasets: 1) the prescription drug plan data in the `Data_main.dta` file; and 2) the subsidy information in the `Data_subsidyinfo.dta` file. Before merging these data, I first create the market share variable in the PDP data as follows:

```
pdp.data <- pdp.data %>%
  group_by(state, year) %>%
  mutate(state_yr_enroll = sum(enrollment, na.rm=TRUE)) %>%
  ungroup() %>%
```

```
  mutate(share = enrollment/state_yr_enroll,
         ln_share = log(share))
```

Then I reshape (or `pivot_longer` in R) the subsidy data so that it matches the format of the PDP data.

```
lis.data <- pivot_longer(lis.data, cols=c("s2006","s2007","s2008","s2009","s2010"),
                         names_to="year",
                         names_prefix="s",
                         values_to="LISsubsidy") %>%
  mutate(year=as.numeric(year))
```

Finally, I merge the two datasets and construct the relevant running variable for the RD analysis (which is the premium net the subsidy). I also construct a couple of flags based on Ericson's code, but I don't end up using these in the analysis:

```
final.data <- pdp.data %>%
  inner_join(lis.data, by=c("PDPregion","year")) %>%
  mutate(LISPremium = premium - LISsubsidy,
         proposedBenchmarkPlan = ifelse(LISPremium<=0,1,0),
         ProblemObs = case_when(
           LISPremium < 0 & LIS == 0 ~ 1,
           LISPremium > 0 & LIS == 1 ~ 2
         ),
         LISPremium = ifelse(benefit=="E",NA,LISPremium),
         proposedBenchmarkPlan = ifelse(benefit=="E", NA, proposedBenchmarkPlan),
         LISPremiumNeg = ifelse(LISPremium<=0, LISPremium, 0),
         LISPremiumPos = ifelse(LISPremium>=0, LISPremium, 0))
```

Question and Answers

Question 1.

Recreate Figure 3 from Ericson (2014).

Figure 1 presents the replication of Ericson's Figure 3. This plot is based on 20 equally spaced bins on each side of the threshold, and we can see a very clear reduction in enrollments among plans just above the threshold. In other words, plans that missed out on the subsidy have much lower enrollments.

Question 2.

Repeat Figure 3 using different bin widths.

Figure 2 presents an alternative plot using 10 equally spaced bins on each side of the threshold. The general takeaways from Figure 3 in Ericson (2014) do not appear to be driven by the selection of the bin width.

Question 3.

Provide the results from the manipulation tests described in Cattaneo et al. (2018).

Such a test is easily implemented in `R` (also in Stata) using the `rdrobust` package. The results from the test are provided in Figure 3.

Question 4.

Summarize the relationship between market shares in 2006 on changes in premiums from 2006 to 2007

There are lots of ways to "summarize a relationship", but I'm just going to present a scatterplot of the changes in premiums and the market shares. This is in Figure 4. The idea behind this graph is to present initial evidence of the "invest-then-harvest" strategy, which is one of the main takeaways from Ericson (2014). He doesn't address this in the same way, but the idea is that high enrollments today might allow plans to increase premiums tomorrow.

Question 5.

Use the Part D low-income subsidy as an IV for market share to examine the effect of market share in 2006 on premium changes.

Again, Ericson (2014) doesn't do this, but to me, a clear identification strategy to estimate the effect of enrollments yesterday on premium changes today would be to simply run that regression. If we're concerned about some unobserved factors influencing market shares and prices, then we can use the low-income subsidy information as an instrument for market share (since we already showed a big discontinuity in Figures 1 and 2).

If we run this simple IV, we get a coefficient estimate on log market share of 3.13 with a 95% confidence interval of 2.74 to 3.52. Since market share is measured in logs and premium changes are measured in dollars, this implies that a doubling of one's market share in period $t-1$ leads to a $3.13 increase in premiums the next year. Considering the mean market share in 2006 is just 2.4%,

the idea of doubling market share seems reasonable, which suggests these are relatively meaningful effects.
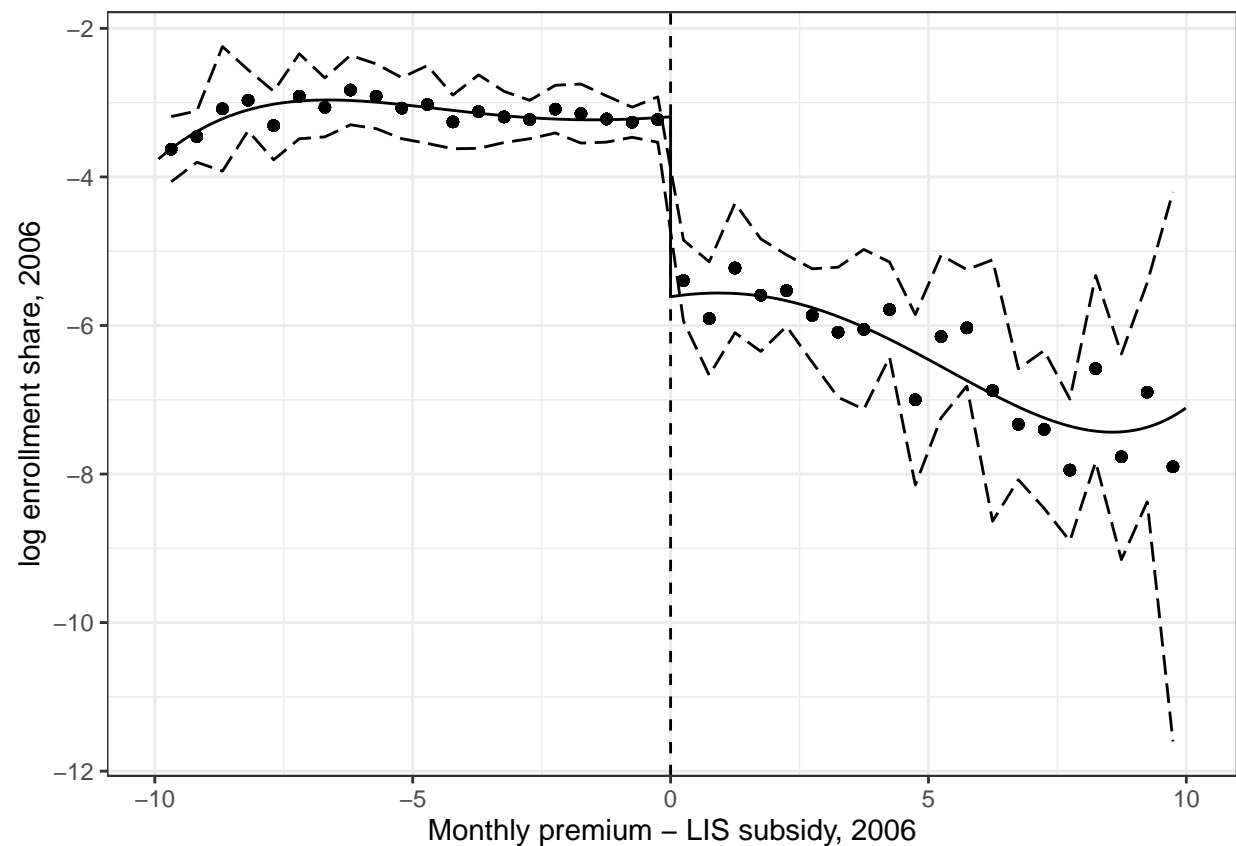
Tables and Figures

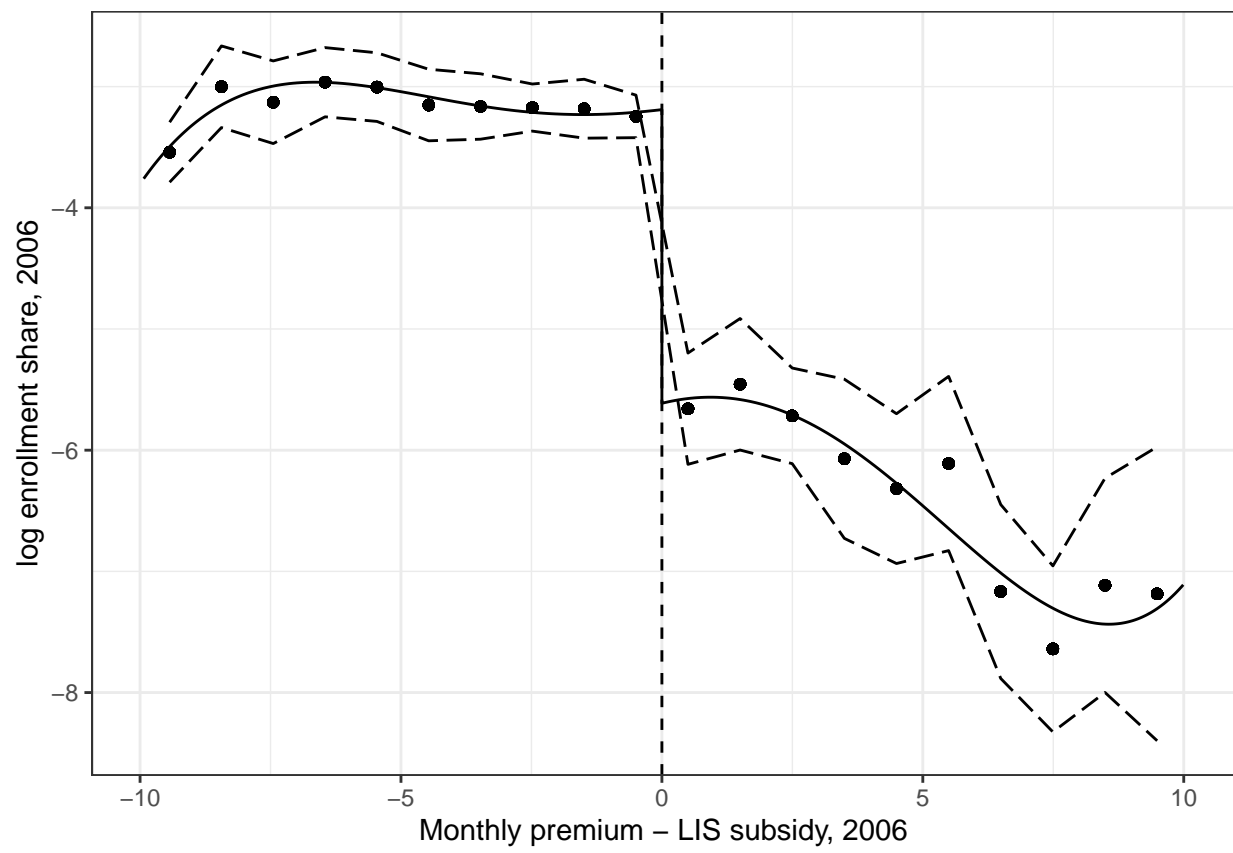Figure 1: Intial Regression Discontinuity Plot

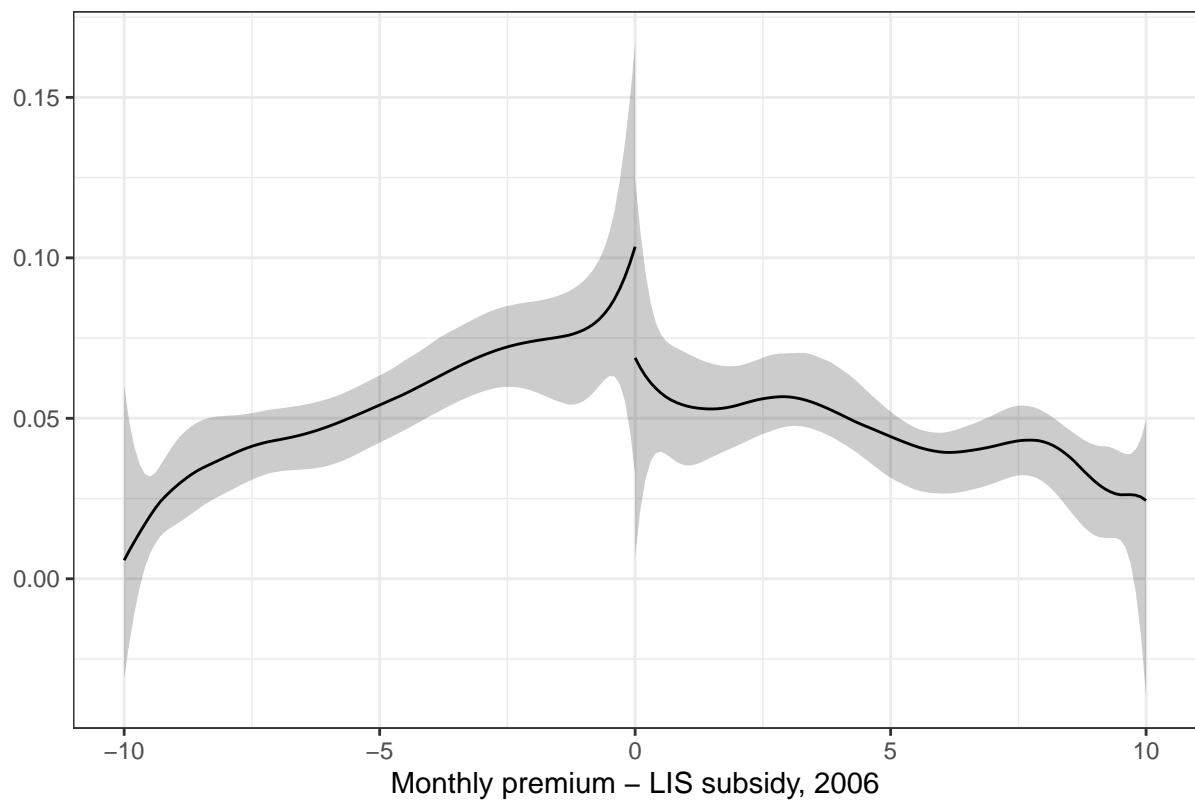Figure 2: Regression Discontinuity Plot with Larger Bins

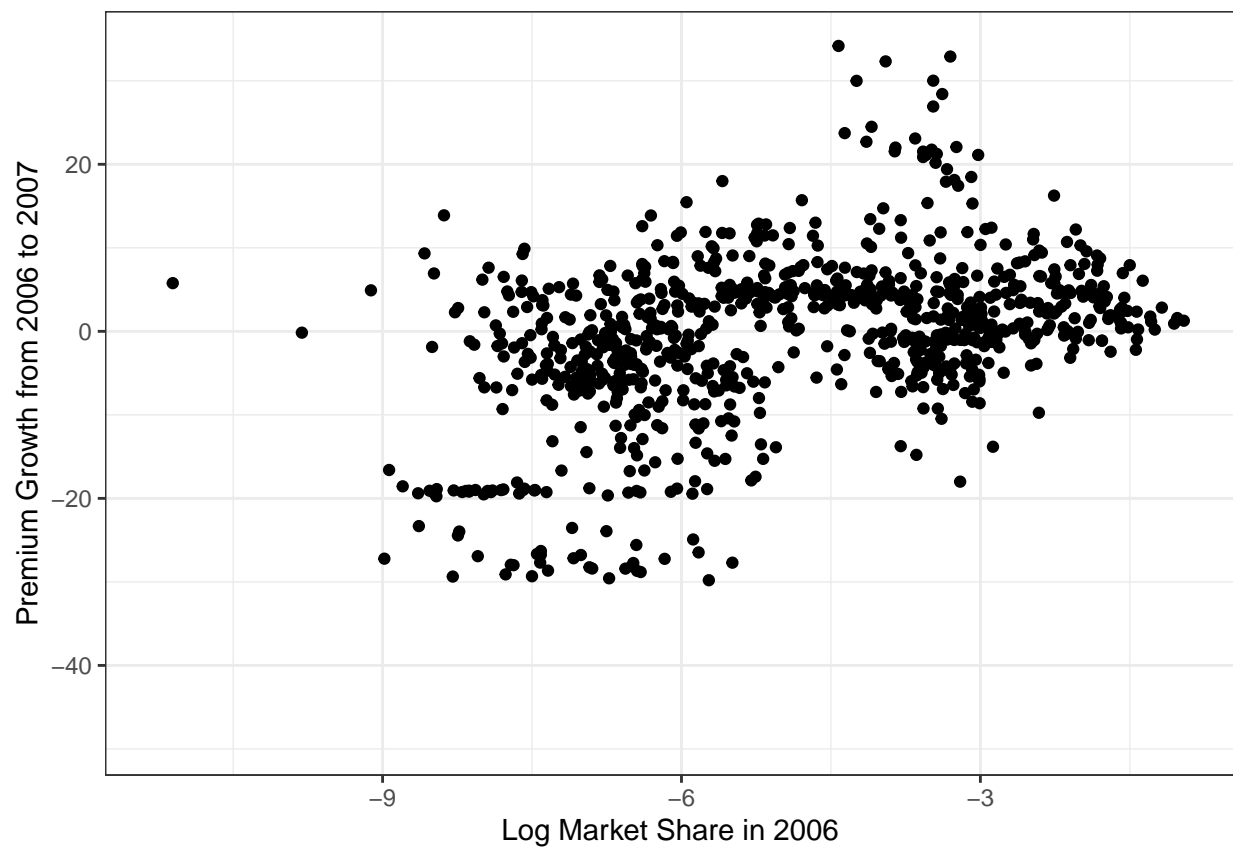Figure 3: Test of Manipulation of Running Variable

Figure 4: Scatterplot of Premium Changes and Market Share

References

Ericson, Keith M. 2014. "Consumer Inertia and Firm Pricing in the Medicare Part d Prescription Drug Insurance Exchange." *American Economic Journal: Economic Policy* 6 (1): 38–64.