Econ771 - Empirical Exercise 3

Nixon Torres Candiales

November 4,2022

Overview

In this assignment, we're going to work through some applied issues related to regression discontinuity designs. We'll cover the basics of strict and fuzzy RD, and we'll work through standard specification tests. We'll also introduce some more technical aspects of bin and bandwidth selection.

Please "submit" your answers as a GitHub repository link on Canvas. In this repo, please include a final document with your main answers and analyses in a PDF. Be sure to include in your repository all of your supporting code files. Practice writing good code and showing me only what I would need to recreate your results.

Resources and data

The data for this assignment comes from the AEJ: Policy website, where Keith Ericson's complete dataset is available. The data are available here. I will also upload the replication files to our class OneDrive folder.

Questions

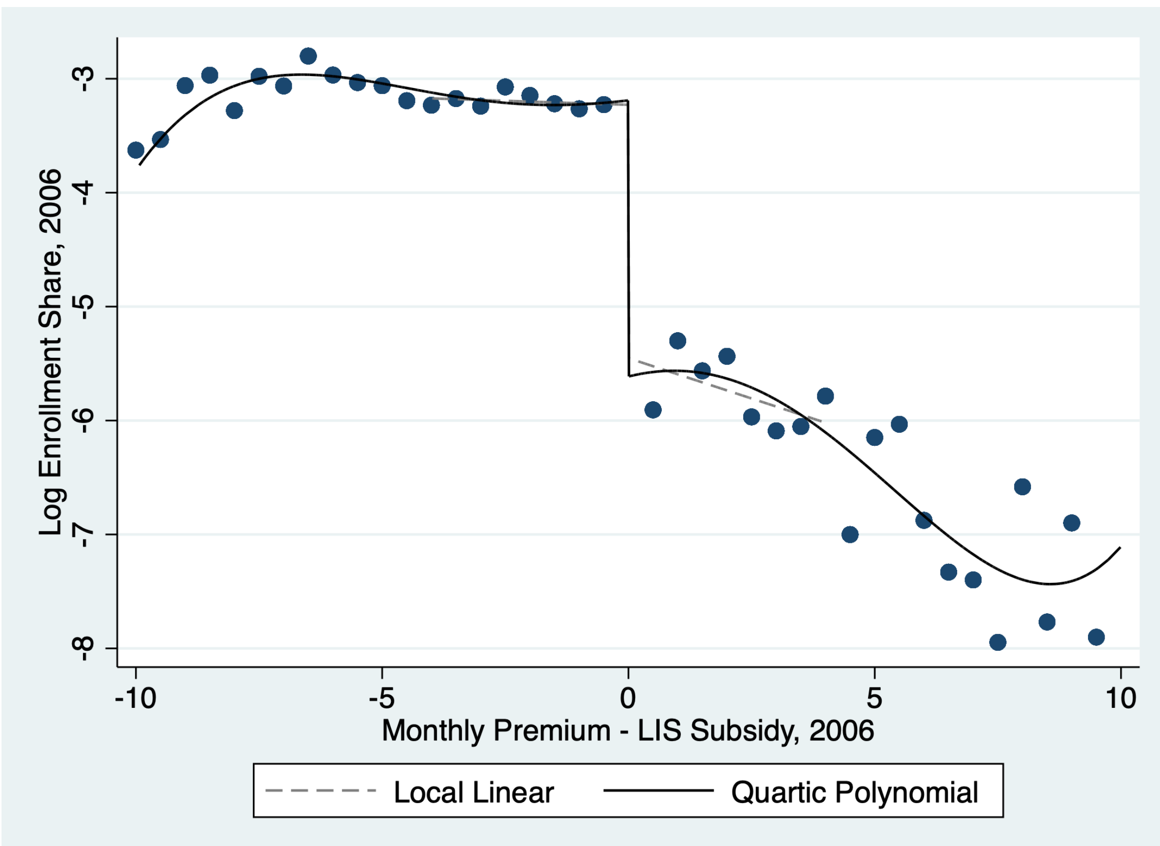
In your GitHub repository, please be sure to clearly address/answer the following questions.

1. Recreate the table of descriptive statistics (Table 1) from @ericson2014.

Table 1: Descriptive Statistics of Medicare Part D Plans

	Cohort (Year of plan introduction)				
	2006	2007	2008	2009	2010
Mean monthly premium	\$37	\$40	\$36	\$30	\$33
	(13)	(17)	(20)	(5)	(9)
Mean deductible	\$92	\$114	\$146	\$253	\$118
	(116)	(128)	(125)	(102)	(139)
Fraction enhanced benefit	0.43	0.43	0.58	0.03	0.69
Fraction of plans offered by firms already offering a plan ...					
... in the United States	0	0.76	0.98	1	0.97
... in the same state	0	0.53	0.91	0.68	0.86
Number of unique firms	51	38	16	5	6
Number of plans	1429	658	202	68	107

2. Recreate Figure 3 from @ericson2014.



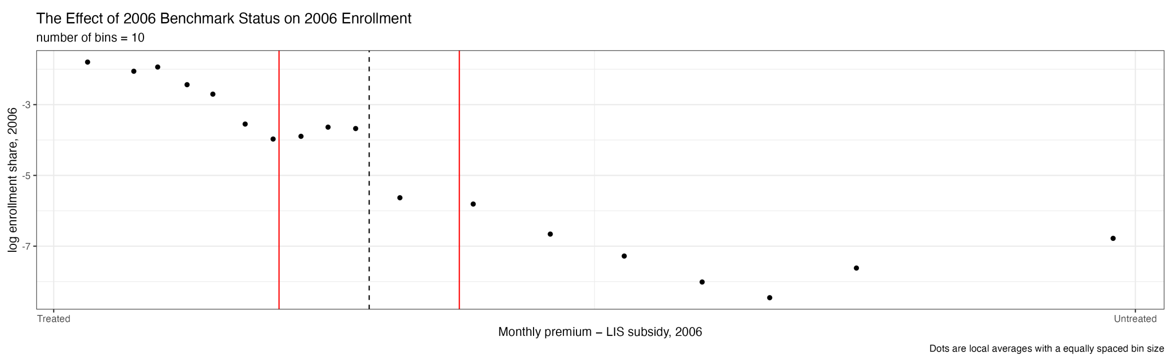
3. @calonico2015 discuss the appropriate partition size for binned scatterplots such as that in Figure 3 of Ericson (2014). More formally, denote by $\mathcal{P}_{-,n} = \{P_{-,j} : j = 1, 2, ...J_{-,n}\}$ and $\mathcal{P}_{+,n} = \{P_{+,j} : j = 1, 2, ...J_{+,n}\}$ the partitions of the support of the running variable x_i on the left and right (respectively) of the cutoff, $\bar{x}$. $P_{-,j}$ and $P_{+,n}$ denote the actual supports for each j partition of size $J_{-,n}$ and $J_{+,n}$, such that $[x_l, \bar{x}) = \bigcup_{j=1}^{J_{-,n}} P_{-,j}$ and $(\bar{x}, x_u] = \bigcup_{j=1}^{J_{+,n}} P_{+,j}$. Individual bins are denoted by $p_{-,j}$ and $p_{+,j}$. With this notation in hand, we can write the partitions $J_{-,n}$ and $J_{+,n}$ with equally-spaced bins as

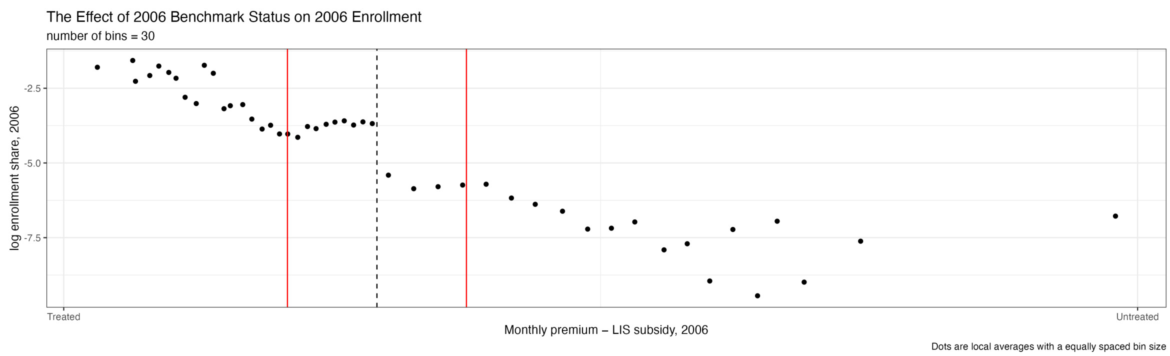
$$p_{-,j} = x_l + j \times \frac{\bar{x} - x_l}{J_{-,n}},$$

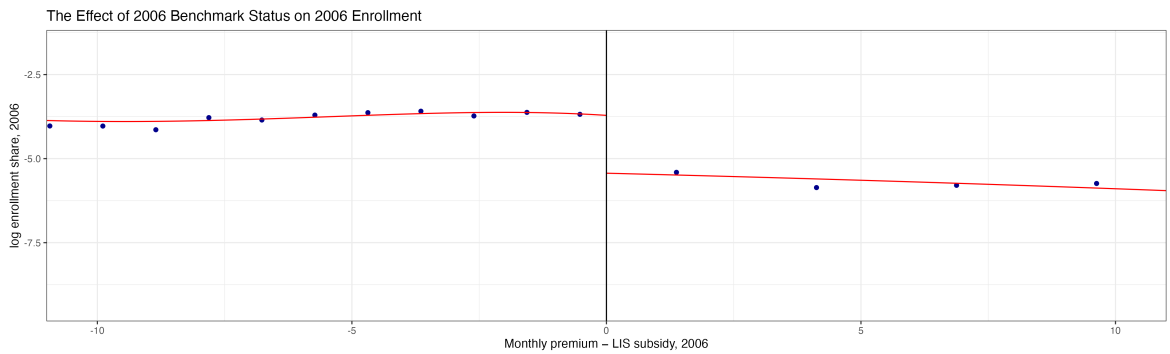
and

$$p_{+,j} = \bar{x} + j \times \frac{x_u - \bar{x}}{J_{+,n}}.$$

Recreate Figure 3 from Ericson (2014) using $J_{-,n} = J_{+,n} = 10$ and $J_{-,n} = J_{+,n} = 30$. Discuss your results and compare them to your figure in Part 2.







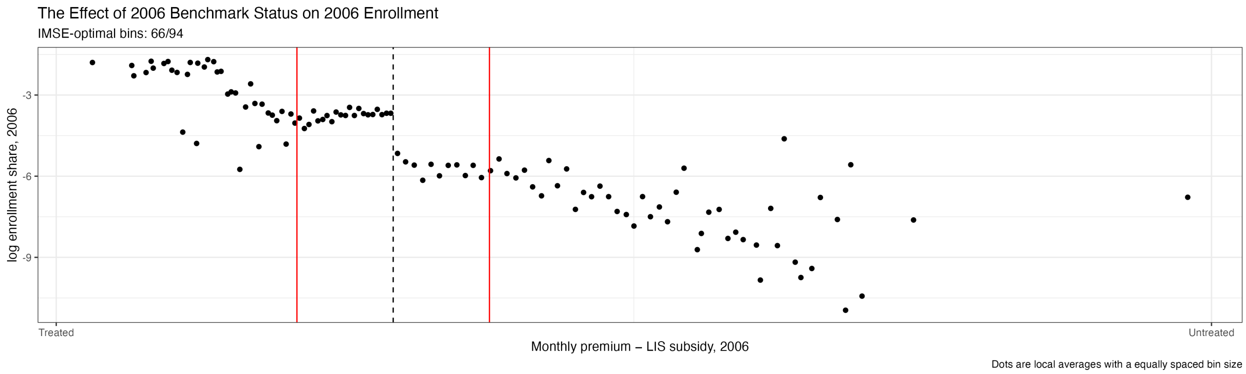
4. With the notation above, @calonico2015 derive the optimal number of partitions for an evenly-spaced (ES) RD plot. They show that

$$J_{ES,-,n} = \left\lceil \frac{V_-}{\mathcal{V}_{ES,-}} \frac{n}{\log(n)^2} \right\rceil$$

and

$$J_{ES,+,n} = \left\lceil \frac{V_+}{\mathcal{V}_{ES,+}} \frac{n}{\log(n)^2} \right\rceil,$$

where V_- and V_+ denote the sample variance of the subsamples to the left and right of the cutoff and $\mathcal{V}_{ES,\cdot}$ is an integrated variance term derived in the paper. Use the `rdrobust` package in `R` (or `Stata` or `Python`) to find the optimal number of bins with an evenly-spaced binning strategy. Report this bin count and recreate your binned scatterplots from parts 2 and 3 based on the optimal bin number.


The Effect of 2006 Benchmark Status on 2006 Enrollment
IMSE-optimal bins: 66/94
log enrollment share, 2006
-3
-6
-9
Treated
Untreated
Monthly premium – LIS subsidy, 2006
Dots are local averages with a equally spaced bin size

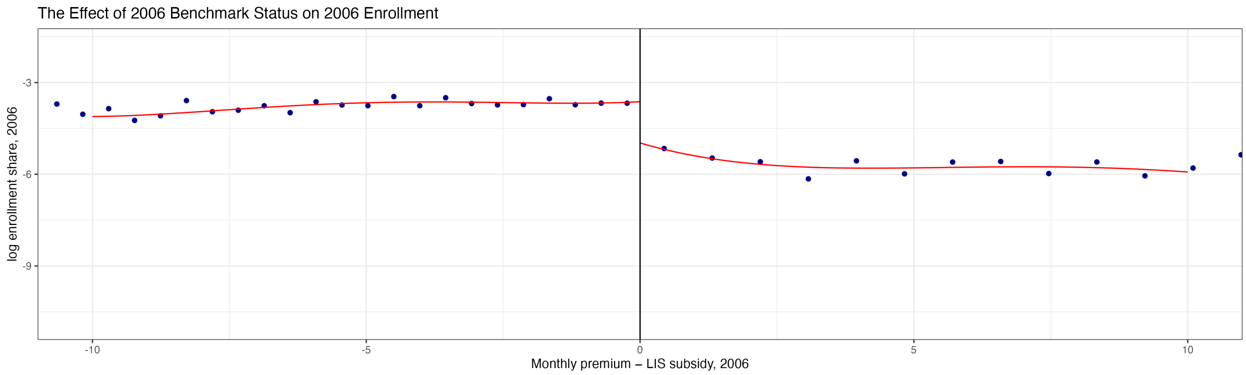

The Effect of 2006 Benchmark Status on 2006 Enrollment
log enrollment share, 2006
-3
-6
-9
-10
-5
0
5
10
Monthly premium – LIS subsidy, 2006

Table 2: Rddensity Test for Different Windows

Window	P-Value
0.03	0.50
0.06	0.42
0.09	0.42
0.12	0.81
0.15	0.74
0.18	0.92
0.21	1.00
0.24	0.86
0.27	0.87
0.30	0.75

5. Although in the online appendix we see evidence of a small discontinuity at the cutoff, after correcting by the optimal bin selection we reject the hypothesis of manipulation in the running variable.

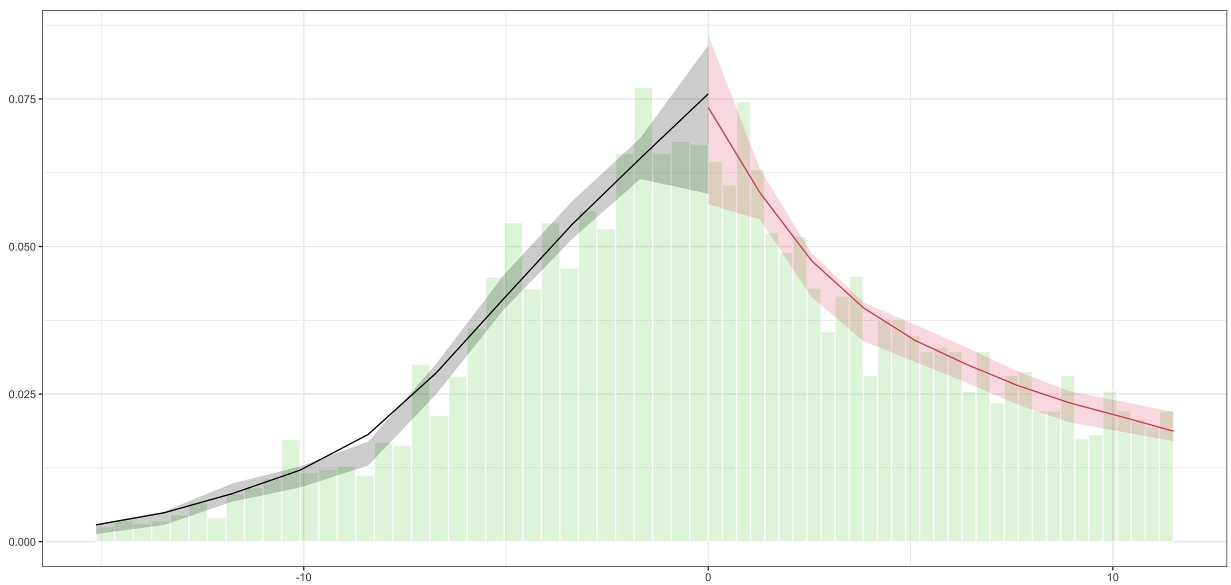


Table 3: Effect of LIS Benchmark Status in 2006 on Plan Enrollment

$\ln s_t$	2006	2007	2008	2009	2010
Panel A. Local linear, bandwidth \$4					
Below benchmark, 2006	2.224***	1.332***	0.902**	0.803*	0.677
	(0.283)	(0.267)	(0.248)	(0.362)	(0.481)
Premium—subsidy, 2006					
Below benchmark	-0.014	-0.077	-0.073	-0.170	-0.215*
	(0.032)	(0.088)	(0.116)	(0.105)	(0.088)
Above benchmark	-0.142+	-0.033	0.049	0.074	0.049
	(0.078)	(0.110)	(0.163)	(0.170)	(0.202)
Num.Obs.	306	299	298	246	212
R2	0.576	0.325	0.131	0.141	0.124
Panel B. Polynomial with controls, bandwidth \$4					
Below benchmark, 2006	2.464***	1.364***	0.872**	0.351	-0.277
	(0.219)	(0.317)	(0.243)	(0.321)	(0.298)
Premium–subsidy, 2006	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
Num.Obs.	306	299	298	246	212
R2	0.794	0.576	0.472	0.535	0.685

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1. Recreate Table 3 of @ericson2014 using the same bandwidth of \$4.00.

We recreate the table as follows, note the standard errors differs at the third decimal. This discrepancy might be due to a different default parameter in R when calculating the standard errors.

7. @calonico2020 show that pre-existing optimal bandwidth calculations (such as those used in @ericson2014) are invalid for appropriate inference. They propose an alternative method to derive minimal coverage error (CE)-optimal bandwidths. Re-estimate your RD results using the CE-optimal bandwidth (`rdrobust` will do this for you) and compare the bandwidth and RD estimates to that in Table 3 of @ericson2014.

Although the coefficients seems to flip signs, this is due to the way the package reports the results. We see how the bin choice increases when the number of observations decreases as well as the bandwidth. The choice of the kernel also affects the optimal bin selection as shown in the table.

Table 4: Rdrobust estimation with optimal bandwith

$\ln s_t$	2006	2007	2008	2009	2010
Panel A. Local linear					
Conventional estimate	-2.29	0.70	0.25	-1.23	-1.07
	(0.55)	(0.69)	(0.48)	(0.59)	(0.88)
Observations	306	245	200	143	128
H	0.75	1.92	2.38	2.12	1.95
Bin	1.59	4.88	7.00	4.89	4.85
Kernel	Uniform	Uniform	Uniform	Uniform	Uniform
Conventional estimate	-2.51	0.96	0.48	-1.20	-1.17
	(0.52)	(0.72)	(0.49)	(0.65)	(0.63)
Observations	306	245	200	143	128
H	0.91	2.24	2.29	2.79	4.41
Bin	1.72	4.49	6.24	5.35	9.67
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
Panel B. Quadratic Polinomial					
Conventional estimate	-2.58	0.77	1.13	-0.67	-0.84
	(0.62)	(0.89)	(0.67)	(0.94)	(1.04)
Observations	306	245	200	143	128
H	1.02	2.61	2.79	2.52	3.20
Bin	2.07	5.36	6.96	5.27	6.13
Kernel	Uniform	Uniform	Uniform	Uniform	Uniform
Conventional estimate	-2.89	0.93	0.92	-0.64	-1.06
	(0.62)	(0.99)	(0.75)	(1.19)	(1.01)
Observations	306	245	200	143	128
H	1.04	2.62	2.77	2.35	4.02
Bin	1.99	5.10	6.02	4.33	6.64
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular

Note: Robust standard erros in parenthesis

Table 5: Effect of LIS Benchmark Status in 2006 on Premiums in Later Year

Premium - Subsidy	2007	2008	2009	2010
Conventional Coef	-0.935	1.267	2.276	0.629
Confidence Interval	[-2.764,1.225]	[-1.927,5.173]	[-0.537,6.863]	[-4.478,8.157]
Observations	569	512	378	322

Note: Replicaton table A.7 Online Appendix. Fuzzy Regression Discontinuity

Table 6: Effect of Market Share in 2006 on Future Premium Changes

	ln(monthly premium)
$\hat{lnS}$	-0.165***
	(0.024)
Num.Obs.	4123
State FE	X
Year FE	X

8. Now let's extend the analysis in Section V of @ericson2014 using IV. Use the presence of Part D low-income subsidy as an IV for market share to examine the effect of market share in 2006 on future premium changes.

Note table 5 presents the estimates for the Fuzzy Regression discontinuity emulating table A7 in the online appendix. Under this setting we find no evidence whether falling above or below the benchmark in 2006 had any effect on average premiums in the subsequent years, as mentioned in Ericson (2014).

9. Discuss your findings and compare results from different binwidths and bandwidths. Compare your results in part 8 to the invest-then-harvest estimates from Table 4 in @ericson2014.

We found similar results than those in table 4 in Ericson (2014). The invest-then-harvest hypothesis seems to explain the price behavior in older plans (on their 5th year). Also, we found that for plans in the second year they seem to decrease in price which can be seen as and "investing" or increasing the market share to in later periods increase prices. Finally, with regards to market share, we found a negative effect, which might be in line with the hypotesis that new plans are priced lower.

10. Reflect on this assignment. What did you find most challenging? What did you find most surprising?

While reading the paper's online appendix I found there was a small discontinuity around the cutoff. However, after testing for manipulation in the forcing variable using the optimal binwidht implemented in rdrobust I found there was no evidence of bunching. This result was nos expected and I need to run further checks and confirm I use the correct sintax on my code to be on the safe there are no code mistakes tat biase my estimates. Overall, the intuition of the RD designs is straight forward.

I still find working with the tables in Rmarkdown challenging but definitely I have seen improvements on my workflow. On the specifics of this assignment, the Fuzzy RD was challenging to implement. Having acces to the paper code and online appendix made the difference in the replication.