Econ771 - Empirical Exercise 3

Nixon Torres Candiales

October 31, 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

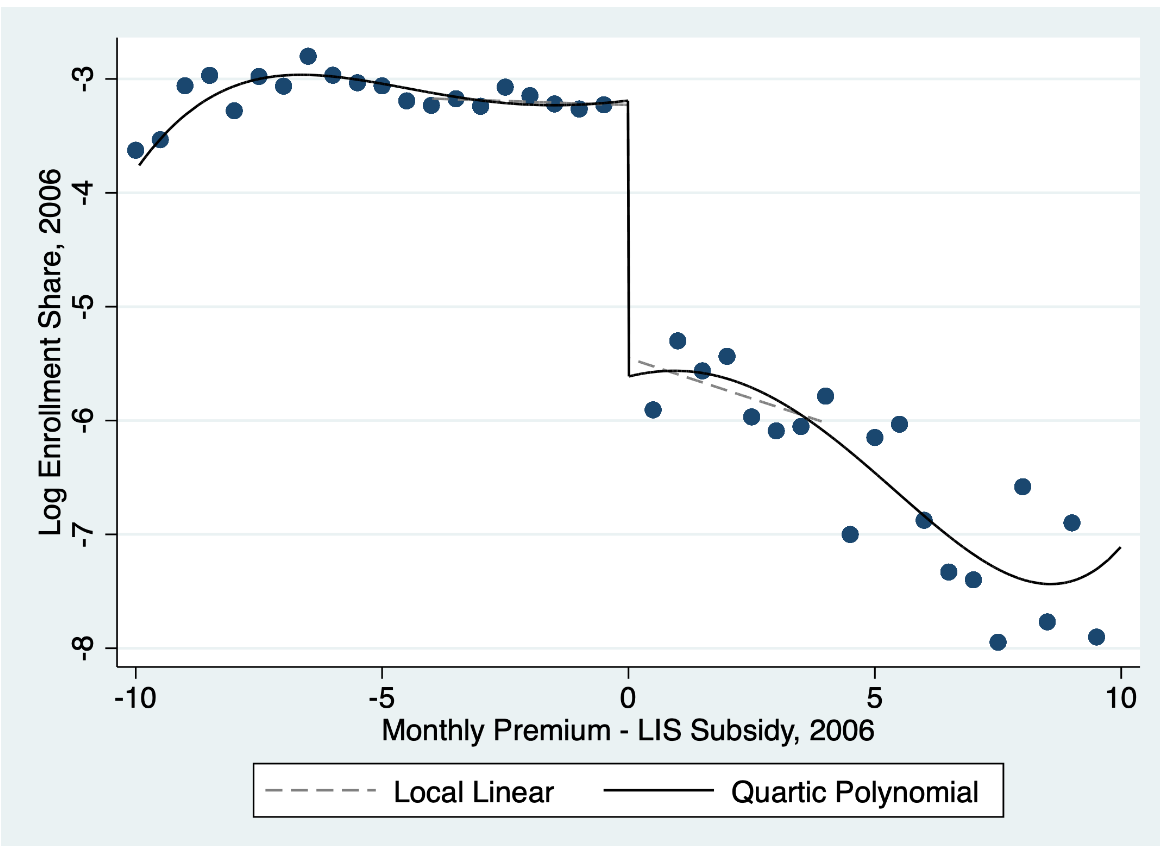
In your GitHub repository, please be sure to clearly address/answer the following questions.

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 $\mathcal{V}_{ES,.}$ 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.

5. One key underlying assumption for RD design is that agents cannot precisely manipulate the running variable. While "precisely" is not very scientific, we can at least test for whether there appears to be a discrete jump in the running variable around the threshold. Evidence of such a jump may suggest that manipulation is present. Provide the results from the manipulation tests described in

    1. This test can be implemented with the `rddensity` package in `R`, `Stata`, or `Python`.

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

Table 2: Effect of LIS Benchmark Status in 2006 on Plan Enrollment

	2006	2007	2008	2009	2010
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
Below benchmark, 2006	2.464***	1.364***	0.872**	0.351	-0.277
	(0.219)	(0.317)	(0.243)	(0.321)	(0.298)
Num.Obs.	306	299	298	246	212

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

1. @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.

2. 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.

3. 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.

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

References