Assignment 3, Health Econ 2
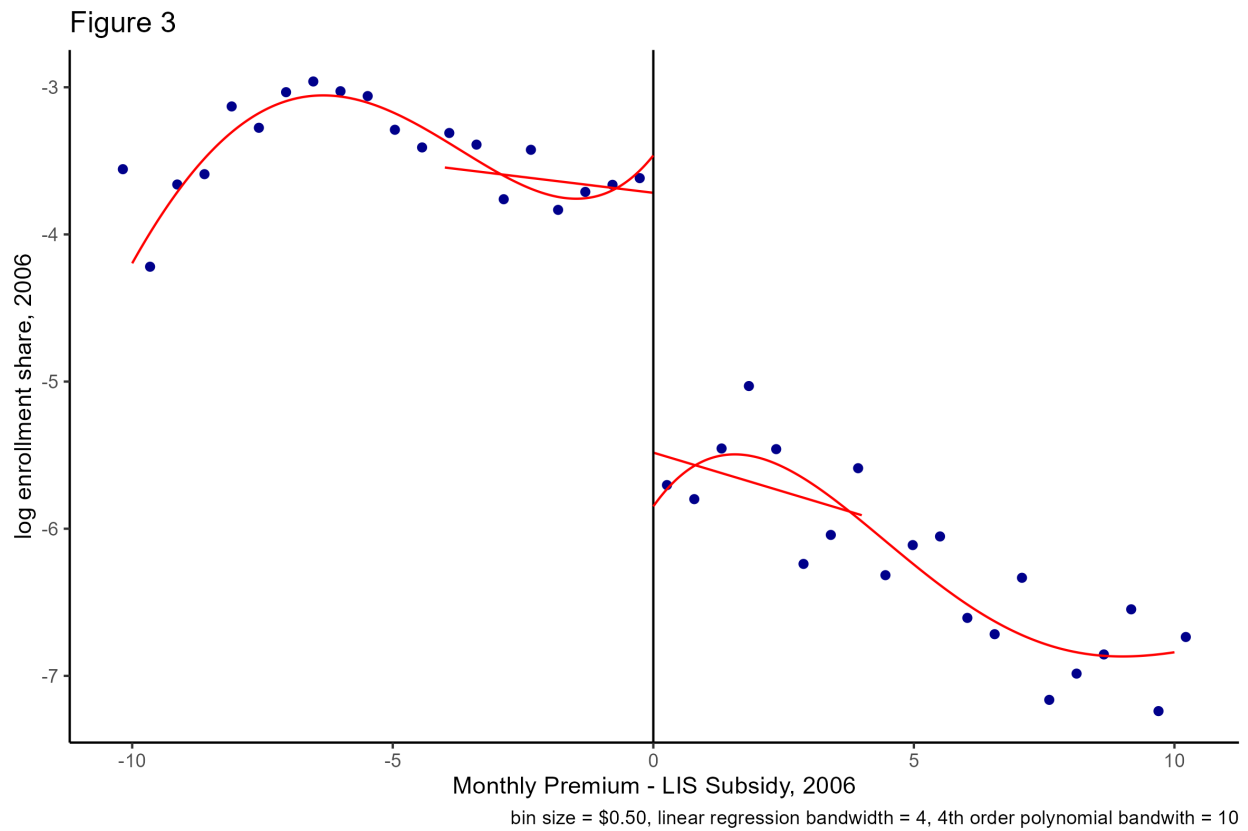
Paul George

2022-11-01

1. Recreate the table of descriptive statistics (Table 1) from Ericson (2014).
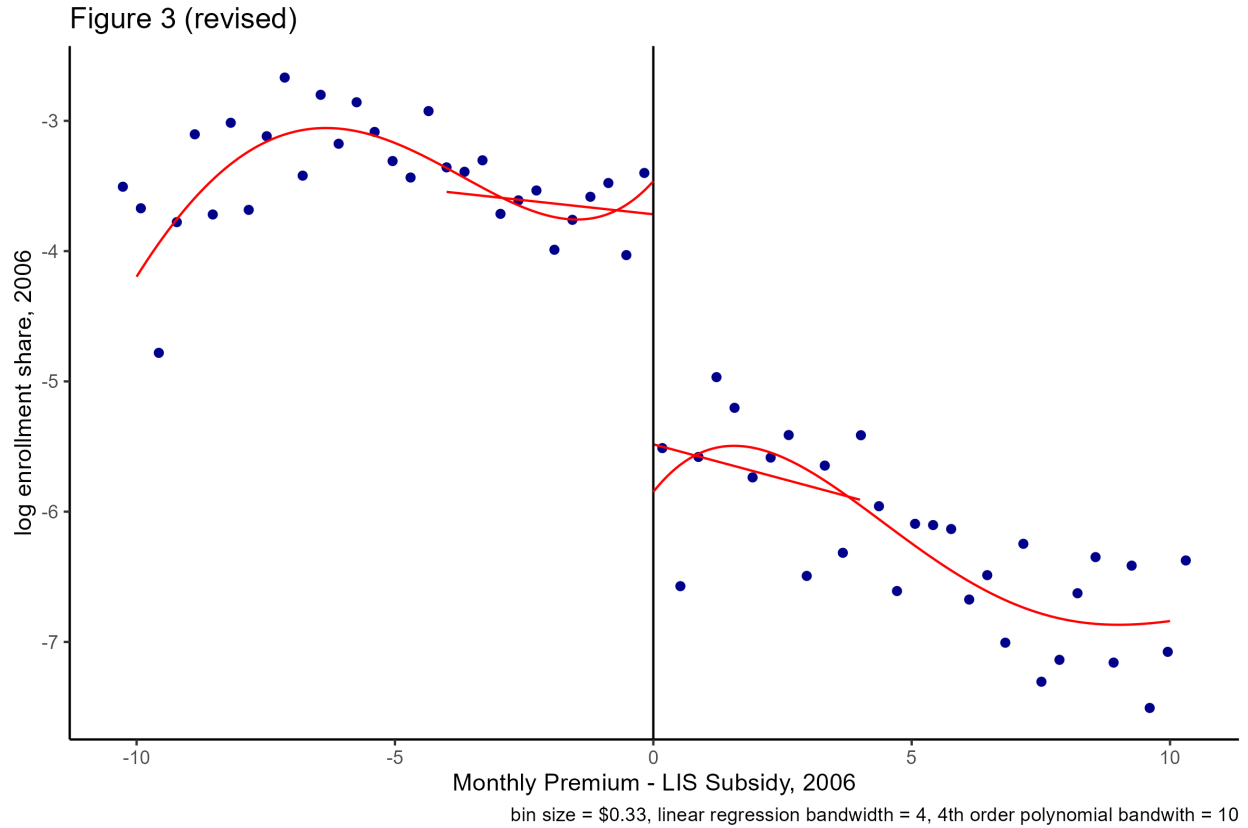
Table 1 - Descriptive Statistics						
variable	subvariable	2006	2007	2008	2009	2010
Mean monthly premium		37	40	36	30	33
standard deviation		13	17	20	5	9
Mean deductible		92	114	146	253	118
standard deviation		116	128	125	102	139
Fraction enhanced benefit		0.43	0.43	0.58	0.03	0.69
Fraction of plans already offered...						
	in the US	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 Ericson (2014).

Figure 3

bin size = $0.50, linear regression bandwidth = 4, 4th order polynomial bandwith = 10

3. Recreate Figure 3 from Ericson (2014). Discuss your results and compare them to your figure in Part 2.

The two figures are actually quite similar, despite the differing number of partitions.

Figure 3 (revised)

log enrollment share, 2006

Monthly Premium - LIS Subsidy, 2006

bin size = $0.33, linear regression bandwidth = 4, 4th order polynomial bandwith = 10

4. 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.

Below, I created bins using IMSE-optimal evenly-spaced method, which returns 7 bins on left, 6 on right.

Figure 4

log enrollment share, 2006

Monthly Premium - LIS Subsidy, 2006

created bins using IMSE-optimal evenly-spaced method, which returns 7 bins on left, 6 on right
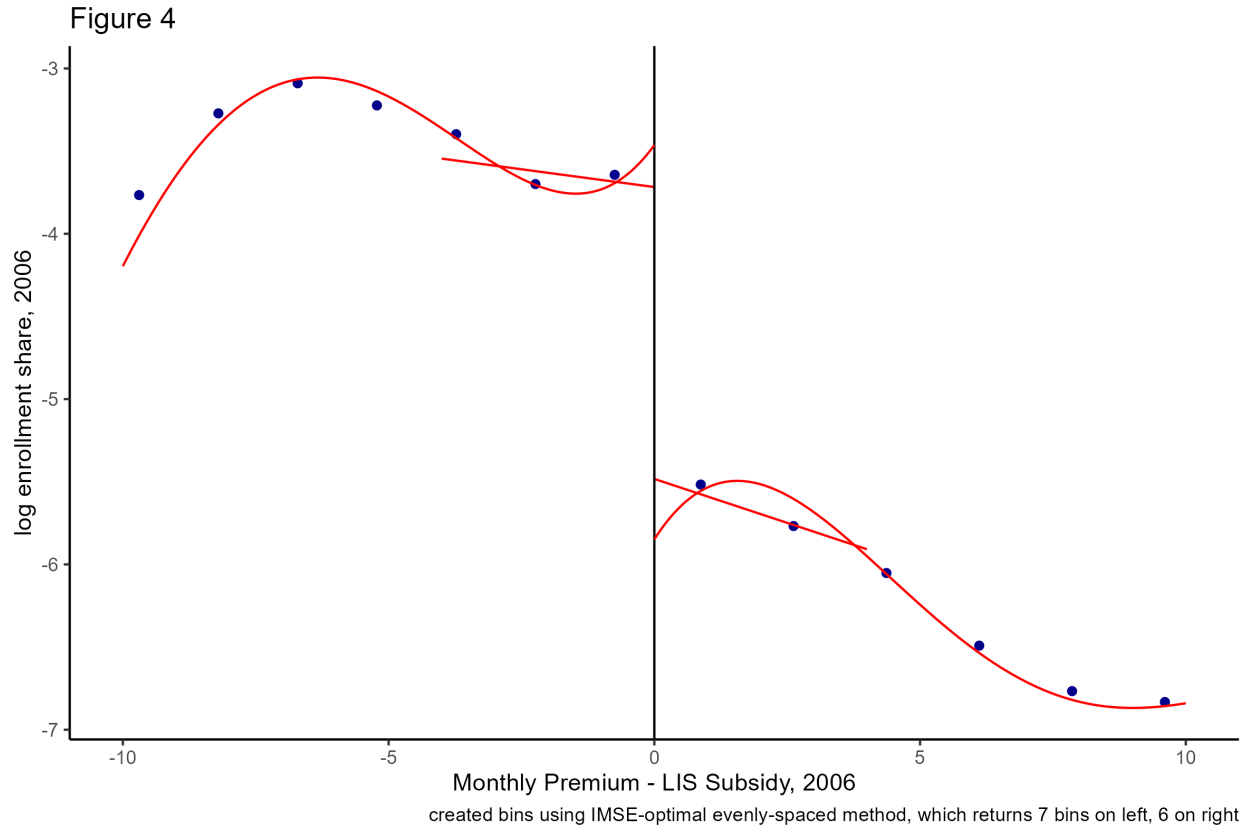
5. Provide the results from the manipulation tests described in Cattaneo, Jansson, and Ma (2018).

Using the above test, we obtain a p value of 0.38, which suggests there is not manipulation around the threshold, which adds evidence to support the regression discontinuity design assumption. Per the paper, this should be expected, as the firms had to choose their premiums prior to the subsidy level announcement. However, this helps in confirming the assumption that firms couldn't manipulate near the RD cutpoint (which would have been theoretically possible, as perhaps firms could have had inside information on what the subsidies would be, and adjusted accordingly).

6. Recreate Table 3 of Ericson (2014) using the same bandwidth of $4.00.

This table below was a doozy to re-create (though I feel like I came up with R code that was able to automatically update the table, which was nice, especially given the next question!). It looks like my column one mostly matches, but then I differ in my results in the subsequent columns. Even comparing my R code to his Stata code, I can't quite figure out what the issue is (though I'm not very familiar with Stata). We have different numbers of observations, which I find odd, because the bin-width inclusion criteria seems straight forward.

Also, of note, I had already created Table 3 by the time we were told to not include Panel C. So I wanted to keep it in, to reflect the work I'd put in!!

variable	subvariable	2006	2007	2008	2009	2010
Table 3 - Recreation						
Panel A - Linear						
Below benchmark 2006	estimate	2.224	2.767	1.437	0.739	0.568
	standard dev	0.244	0.135	0.158	0.144	0.158
	p value	0	0	0	0	0
Premium - subsidy, 2006						
	below estimate	-0.014	-0.075	-0.136	-0.358	-0.224
	standard dev	0.073	0.057	0.075	0.068	0.075
	p value	0.847	0.191	0.069	0	0.003
	above estimate	-0.142	-0.196	-0.136	-0.348	-0.364
	standard dev	0.081	0.067	0.074	0.065	0.08
	p value	0.08	0.004	0.065	0	0
obs		306	328	251	175	147
R2		0.576	0.604	0.312	0.462	0.336
Panel B - Polynomial						
Below benchmark, 2006	estimate	2.491	2.043	1.319	0.949	0.519
	standard dev	0.305	0.135	0.137	0.184	0.179
	p value	0	0	0	0	0.005
Premium - subsidy, 2006						
obs		306	328	251	175	147
R2		0.792	0.881	0.862	0.801	0.81
Panel C - Interactions						
Below benchmark or de min						
	2006 + Current year	2.491	2.799	1.107	1.945	2.486
		0.305				
		0				
	2006 and not current year		2.318	1.577	0.984	0.662
	current year but not 2006		0.996	0.02	1.102	2.043
Premium - subsidy, 2006						
obs		306	328	251	175	147
R2		0.792	0.889	0.867	0.822	0.855

7. 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 Ericson (2014).

Note from the package documentation, h specifies the main bandwidth, and b specifies the bias bandwidth. Thus, we obtain our bandwidth per the specification should be 3.05 and 3.05, and the table below displays those results. Again, these results are quite similar to the previous 2 tables. I can't really find a pattern, some of the estimates are a bit higher, others are a bit lower. My a priori guess would have been that the smaller the bandwidth, the higher the estimates in general, but that doesn't seem to be the case.

variable	subvariable	2006	2007	2008	2009	2010
Table 3 - New bandwidth						
Panel A - Linear						
Below benchmark 2006	estimate	2.298	2.987	1.234	0.684	0.505
	standard dev	0.256	0.154	0.188	0.166	0.17
	p value	0	0	0	0	0.004
Premium - subsidy, 2006						
	below estimate	-0.027	-0.159	-0.352	-0.385	-0.105
	standard dev	0.102	0.094	0.131	0.112	0.112
	p value	0.792	0.093	0.008	0.001	0.349
	above estimate	-0.046	-0.134	-0.09	-0.572	-0.675
	standard dev	0.114	0.104	0.136	0.107	0.12
	p value	0.687	0.198	0.51	0	0
obs		230	253	182	136	116
R2		0.591	0.626	0.263	0.439	0.393
Panel B - Polynomial						
Below benchmark, 2006	estimate	2.32	2.137	1.142	0.878	0.366
	standard dev	0.362	0.151	0.187	0.205	0.24
	p value	0	0	0	0	0.133
Premium - subsidy, 2006						
obs		230	253	182	136	116
R2		0.803	0.905	0.864	0.827	0.82
Panel C - Interactions						
Below benchmark or de min						
	2006 + Current year	2.491	2.972	0.798	1.918	2.739
		0.305				
		0				
	2006 and not current year		2.398	1.576	0.907	0.668
	current year but not 2006		1.104	-0.088	1.125	2.309
Premium - subsidy, 2006						
obs		306	253	182	136	116
R2		0.792	0.912	0.873	0.845	0.856

8. Now let's extend the analysis in Section V of Ericson (2014) 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.

Table 5 - IV model

Low Income Subsidy in 2006 as instrument for enrollment share in 2006

	IV
Instrument	-0.073
	(0.026)
Year of Plan 2	-0.066
	(0.070)
Year of Plan 3	0.006
	(0.090)
Year of Plan 4	0.181
	(0.093)
Year of Plan 5	0.271
	(0.104)

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 Ericson (2014).

As discussed above, the binwidths and bandwidths did not affect the estimates very much. I think this adds robustness to the model, because it could be problematic if the model and model's significance and findings were very dependent on the binwidth (especially binwidth) or bandwidth.

The results from question 8 are similar to the results from Ericson Table 4, in that they show an increasing effect as the year of plan increases.

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

I really enjoyed this assignment overall - I think RD estimators, when applicable, are pretty straight-forward and much more understandable than IV, especially to an audience that's unfamiliar with causal inference (such as the audience for most medical literature). I plan to use RD in the future, so I found this exercise particularly helpful. I struggled a bit at first coming up with the plots (the most important part!), but once I got a handle on rdrobust package, it's actually pretty straight-forward.