Regression Discontinuity

Christopher Taber

Department of Economics
University of Wisconsin-Madison

February 28, 2018

I will describe the basic ideas of RD, but ignore many of the details

Good references (and things I used in preparing this are):

- "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," Hahn, Todd, and Van der Klaauw, EMA (2001)
- "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," McCrary, Journal of Econometrics (2008)
- "Regression Discontinuity Designs: A Guide to Practice," Imbens and Lemieux, Journal of Econometrics (2008)
- "Regression Discontinuity Designs in Economics," Lee and Lemiux, JEL (2010)

You can also find various Handbook chapters or Mostly Harmless Econometrics which will help as well

The idea of regression discontinuity goes way back, but it has gained in popularity in recent years

The basic idea is to recognize that in many circumstances policy rules vary at some cutoff point

To think of the simplest case suppose the treatment assignment rule is:

$$T_i = \begin{cases} 0 & X_i < x^* \\ 1 & X_i \geq x^* \end{cases}$$

Many different rules work like this.

Examples:

- Whether you pass a test
- Whether you are eligible for a program
- Who wins an election
- Which school district you reside in
- Whether some punishment strategy is enacted
- Birth date for entering kindergarten

This last one should look pretty familiar-Angrist and Krueger's quarter of birth was essentially a regression discontinuity design

The key insight is that right around the cutoff we can think of people slightly above as identical to people slightly below

Formally we can write it the model as:

$$Y_i = \alpha T_i + \varepsilon_i$$

If

$$E(\varepsilon_i \mid X_i = x)$$

is continuous then the model is identified (actually all you really need is that it is continuous at $x = x^*$)

To see it is identified not that

$$lim_{x \uparrow x^*} E(Y_i \mid X_i = x) = E(\varepsilon_i \mid X_i = x^*)$$
$$lim_{x \downarrow x^*} E(Y_i \mid X_i = x) = \alpha + E(\varepsilon_i \mid X_i = x^*)$$

Thus

$$\alpha = lim_{x \downarrow x^*} E(Y_i \mid X_i = x) - lim_{x \uparrow x^*} E(Y_i \mid X_i = x)$$

Thats it

What I have described thus far is referred to as a “Sharp Regression Discontinuity”

There is also something called a “Fuzzy Regression Discontinuity”

This occurs when rules are not strictly enforced

Examples

- Birth date to start school
- Eligibility for a program has other criterion
- Whether punishment kicks in (might be an appeal process)

This isn't a problem as long as

$$lim_{x \uparrow x^*} E(T_i \mid X_i = x) > lim_{x \downarrow x^*} E(T_i \mid X_i = x)$$

To see identification we now have

$$\begin{aligned}
&\frac{lim_{x \uparrow x^*} E(Y_i \mid X_i = x) - lim_{x \downarrow x^*} E(Y_i \mid X_i = x)}{lim_{x \uparrow x^*} E(T_i \mid X_i = x) - lim_{x \downarrow x^*} E(T_i \mid X_i = x)} \\
&= \frac{\alpha \left[lim_{x \uparrow x^*} E(T_i \mid X_i = x) - lim_{x \downarrow x^*} E(T_i \mid X_i = x) \right]}{lim_{x \uparrow x^*} E(T_i \mid X_i = x) - lim_{x \downarrow x^*} E(T_i \mid X_i = x)} \\
&= \alpha
\end{aligned}$$

Note that this is essentially just Instrumental variables (this is often referred to as the Wald Estimator)

You can also see that this works when T_i is continuous

How do we do this in practice?

There are really two approaches.

The first comes from the basic idea of identification, we want to look directly to the right and directly to the left of the policy change

Lets focus on the Sharp case-we can get the fuzzy case by just applying to Y_i and T_i and then taking the ratio

The data should look something like this (in stata)

We can think about estimating the end of the red line and the end of the green line and taking the difference

This is basically just a version of nonparametric regression at these two points

Our favorite way to estimate nonparametric regression in economics is by Kernel regression

Let $K(x)$ be a kernel that is positive and non increasing in $|x|$ and is zero when $|x|$ is large

Examples:

- Normal pdf: $exp(-x^2)$ Normal
- Absolute value: Absolute

$$\begin{cases} 1-|x| & |x|<1 \\ 0 & |x|\geq 1 \end{cases}$$

- Uniform: $1(|x|<1)$ Uniform
- Epanechnikov kernel: Epanechnikov

$$\begin{cases} \frac{3}{4}\left(1-u^2\right) & |x|<1 \\ 0 & |x|\geq 1 \end{cases}$$

The kernel regressor is defined as

$$E\left(Y \mid X = x\right) \approx \frac{\sum_{i=1}^{N} K\left(\frac{X_i - x}{h}\right) Y_i}{\sum_{i=1}^{N} K\left(\frac{X_i - x}{h}\right)}$$

where h is the bandwidth parameter

Note that this is just a weighted average

- it puts higher weight on observations closer to x
- when h is really big we put equal weight on all observations
- when h is really small, only the observations that are very close to x influence it

This is easiest to think about with the uniform kernel

In this case

$$K\left(\frac{X_i - x}{h}\right) = 1(|X_i - x| < h)$$

So we use take a simple sample mean of observations within h units of X_i

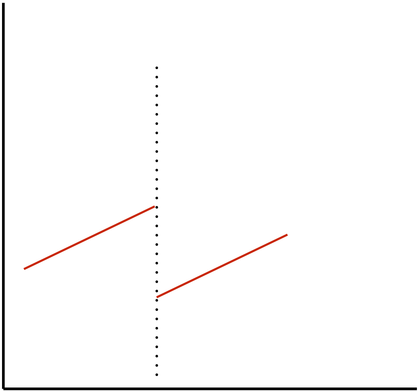
Clearly in this case as with other kernels, as the sample size goes up, h goes down so that asymptotically we are only putting weight on observations very close to x

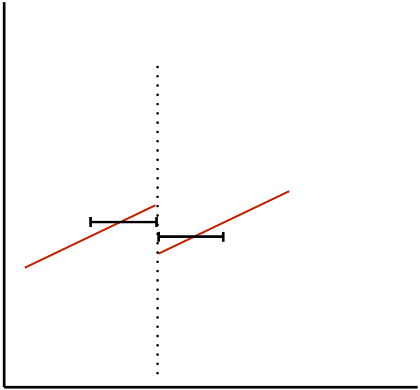
To estimate $lim_{x \downarrow x^*} E(T_i \mid X_i = x)$ we only want to use values of X_i to the right of x^*, so we would use

$$lim_{x \downarrow x^*} E(T_i \mid X_i = x) \approx \frac{\sum_{i=1}^{N} 1\left(X_i > x^*\right) K(\frac{X_i - x^*}{h}) Y_i}{\sum_{i=1}^{N} 1\left(X_i > x^*\right) K(\frac{X_i - x^*}{h})}$$

However it turns out that this has really bad properties because we are looking at the end point

For example suppose the data looked like this





For any finite bandwidth the estimator would be biased downward

It is better to use local linear (or polynomial) regression.

Here we choose

$$\left(\hat{a}, \hat{b}\right) = argmin_{a,b} \sum_{i=1}^{N} K\left(\frac{X_i - x^*}{h}\right) \left[Y_i - a - b(X_i - x^*)\right]^2 1\left(X_i \geq x^*\right)$$

Then the estimate of the right hand side is $\widehat{a}$.

We do the analogous thing on the other side:

$$\left(\hat{a}, \hat{b}\right) = argmin_{a,b} \sum_{i=1}^{N} K\left(\frac{X_i - x^*}{h}\right) \left[Y_i - a - b(X_i - x^*)\right]^2 1\left(X_i < x^*\right)$$

(which with a uniform kernel just means running a regression using the observations between $x^* - h$ and x^*

Lets try this in stata

There is another approach to estimating the model

Define

$$g(x) = E(\varepsilon_i \mid X_i = x)$$

then

$$E(Y_i \mid X_i, T_i) = \alpha T_i + g(X_i)$$

where g is a smooth function

Thus we can estimate the model by writing down a smooth flexible functional form for g and just estimate this by OLS

The most obvious functional form that people use is a polynomial

There are really two different ways to do it:

$$Y_i = \alpha T_i + b_0 + b_1 X_i + b_2 X_i^2 + v_i$$

or

$$\begin{aligned} Y_i = &\alpha T_i + b_0 + b_1 X_i 1\left(X_i < x\right) + b_2 X_i^2 1\left(X_i < x\right) \\ &+ b_3 X_i 1\left(X_i \geq x\right) + b_4 X_i^2 1\left(X_i \geq x\right) + v_i \end{aligned}$$

Lee and Lemieux say the second is better

Note that this is just as “nonparametric” as the Kernel approach

- You must promise to increase the degree of the polynomial as you increase the sample size (in the same way that you lower the bandwidth with the sample size)
- You still have a practical problem of how to choose the degree of the polynomial (in the same way you have a choice about how to choose the bandwidth in the kernel approaches)

You can do both and use a local polynomial-in one case you promise to lower the bandwidth, in the other you promise to add more terms, you could do both

Also, for the “fuzzy” design we can just do IV

Problems

While RD is often really nice, there are three major problems that arise

The first is kind of obvious from what we are doing-and is an estimation problem rather than an identification problem

Often the sample size is not very big and as a practical matter the bandwidth is so large (or the degree of the polynomial so small) that it isn't really regression discontinuity that is identifying things

The second problem is that there may be other rules changes happening at the same cutoff so you aren't sure what exactly you are identifying

One suggestion to test for this is to look at observable characteristics

The third is if the running variable is endogenous

Clearly if people choose X_i precisely the whole thing doesn't work

For example suppose

- carrying 1 pound of drugs was a felony, but less than 1 was a misdemeanor
- people who get their paper in by 5:00 on thursday afternoon are on time, 5:01 is late and marked down by a grade

Note that you need X_i to be precisely manipulated, if there is still some randomness on the actual value of X_i, rd looks fine

Mccrary (2008) suggests to test for this by looking at the density around the cutoff point:

- Under the null the density should be continuous at the cutoff point
- Under the alternative, the density would increase at the kink point when T_i is viewed as a good thing

Lets look at some examples

Randomized Experiments from Non-random Selection in U.S. House Elections

Lee, Journal of Econometrics, 2008

One of the main points of this paper is that the running variable can be endogenous as long as it can not be perfectly chosen.

In particular it could be that:

$$X_i = W_i + \xi_i$$

where W_i is chosen by someone, but ξ_i is random and unknown when W_i is chosen

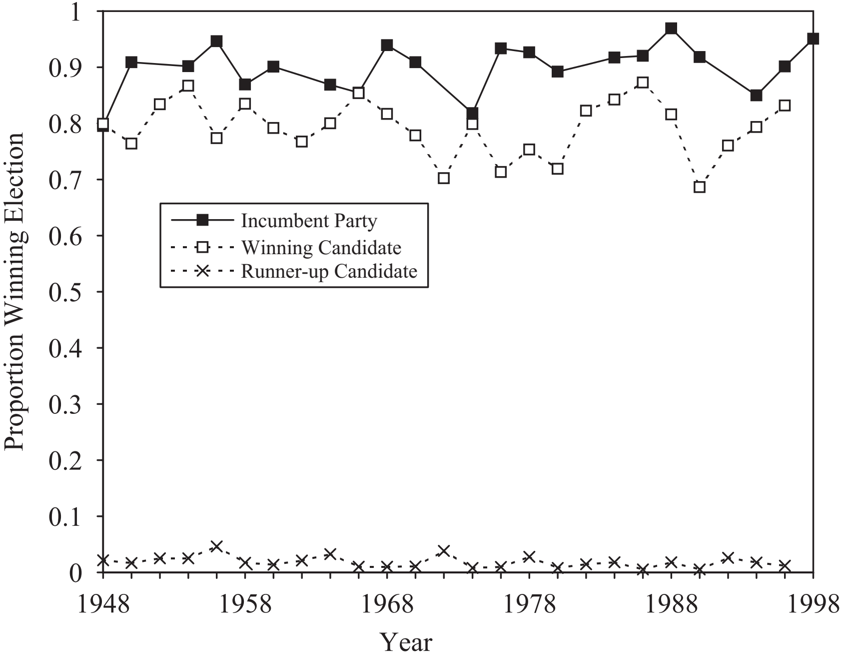
Lee shows that regression discontinuity approaches still work in this case

Incumbency

We can see that incumbents in congress are re-elected at very high rates

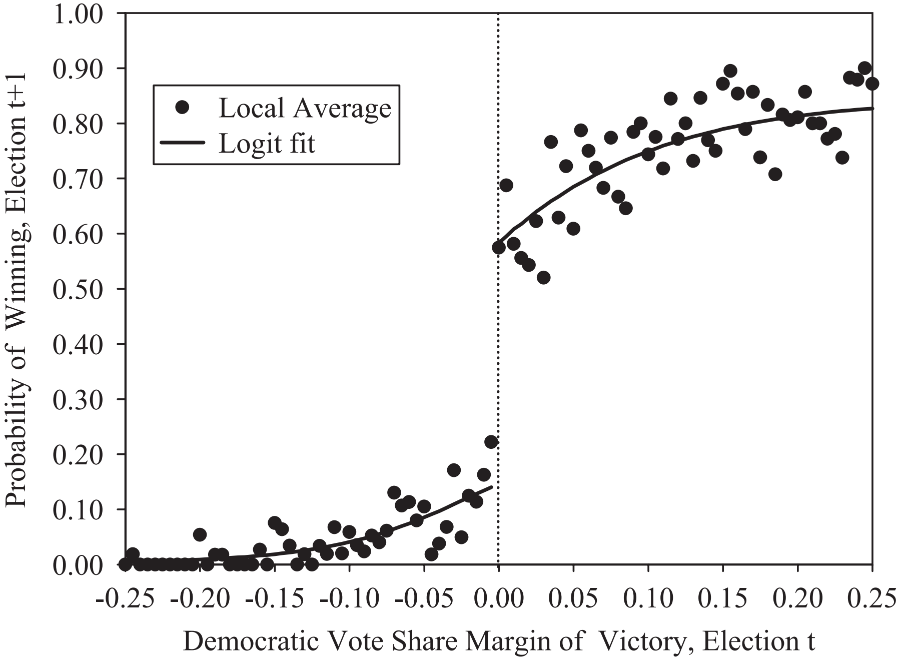
Is this because there is an effect of incumbency or just because of serial correlation in preferences?

Regression discontinuity helps solves this problem-look at people who just barely won (or lost).



Proportion Winning Election
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1948
1958
1968
1978
1988
1998
Year
Incumbent Party
Winning Candidate
Runner-up Candidate

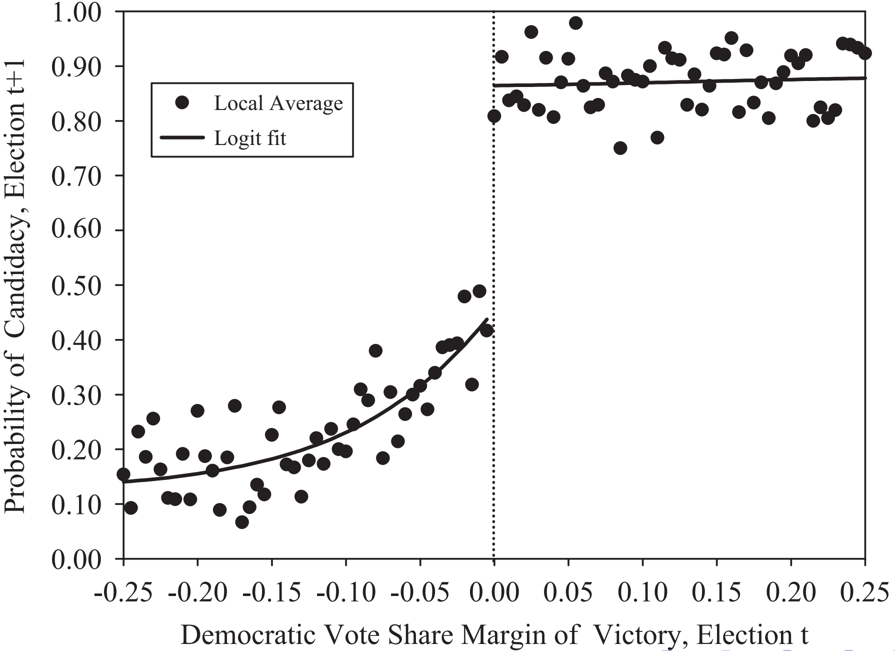
a



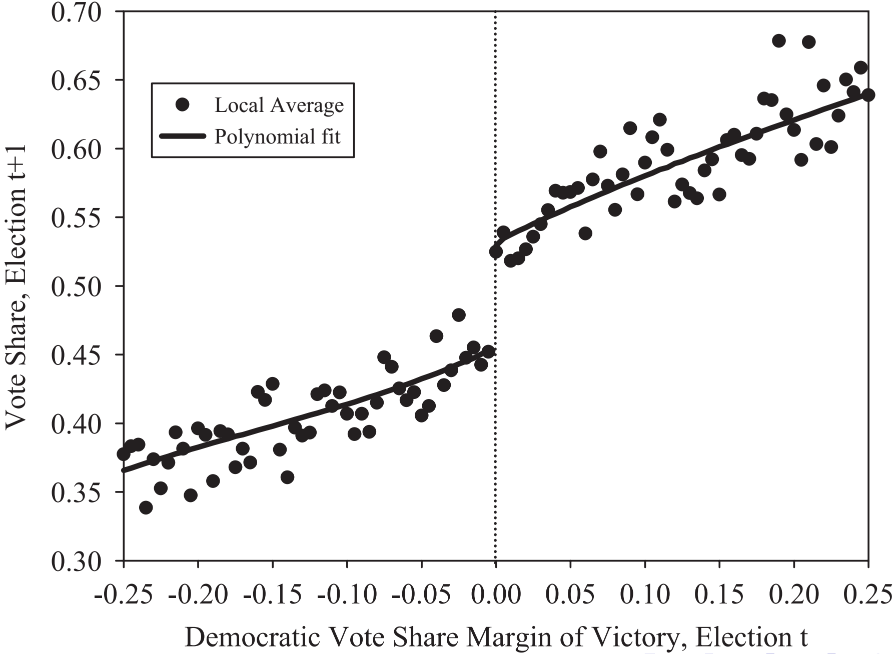
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Probability of Winning, Election t+1
Local Average
Logit fit
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
Democratic Vote Share Margin of Victory, Election t

b

a


Local Average
Logit fit
Probability of Candidacy, Election t+1
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
Democratic Vote Share Margin of Victory, Election t

a



0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
Vote Share, Election t+1
Local Average
Polynomial fit
-0.25 -0.20 -0.15 -0.10 -0.05 0.00 0.05 0.10 0.15 0.20 0.25
Democratic Vote Share Margin of Victory, Election t

Table 1
Electoral outcomes and pre-determined election characteristics: democratic candidates, winners vs. losers: 1948–1996

Variable	All		|Margin| < .5		|Margin| < .05		Parametric fit	
	Winner	Loser	Winner	Loser	Winner	Loser	Winner	Loser
Democrat vote share election $t+1$	0.698	0.347	0.629	0.372	0.542	0.446	0.531	0.454
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.008)	(0.008)
	[0.179]	[0.15]	[0.145]	[0.124]	[0.116]	[0.107]		
Democrat win prob. election $t+1$	0.909	0.094	0.878	0.100	0.681	0.202	0.611	0.253
	(0.004)	(0.005)	(0.006)	(0.006)	(0.026)	(0.023)	(0.039)	(0.035)
	[0.276]	[0.285]	[0.315]	[0.294]	[0.458]	[0.396]		
Democrat vote share election $t-1$	0.681	0.368	0.607	0.391	0.501	0.474	0.477	0.481
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.008)	(0.009)	(0.01)
	[0.189]	[0.153]	[0.152]	[0.129]	[0.129]	[0.133]		
Democrat win prob. election $t-1$	0.889	0.109	0.842	0.118	0.501	0.365	0.419	0.416
	(0.005)	(0.006)	(0.007)	(0.007)	(0.027)	(0.028)	(0.038)	(0.039)
	[0.31]	[0.306]	[0.36]	[0.317]	[0.493]	[0.475]		
Democrat political experience	3.812	0.261	3.550	0.304	1.658	0.986	1.219	1.183
	(0.061)	(0.025)	(0.074)	(0.029)	(0.165)	(0.124)	(0.229)	(0.145)
	[3.766]	[1.293]	[3.746]	[1.39]	[2.969]	[2.111]		
Opposition political experience	0.245	2.876	0.350	2.808	1.183	1.345	1.424	1.293
	(0.018)	(0.054)	(0.025)	(0.057)	(0.118)	(0.115)	(0.131)	(0.17)
	[1.084]	[2.802]	[1.262]	[2.775]	[2.122]	[1.949]		
Democrat electoral experience	3.945	0.464	3.727	0.527	1.949	1.275	1.485	1.470
	(0.061)	(0.028)	(0.075)	(0.032)	(0.166)	(0.131)	(0.23)	(0.151)
	[3.787]	[1.457]	[3.773]	[1.55]	[2.986]	[2.224]		
Opposition electoral experience	0.400	3.007	0.528	2.943	1.375	1.529	1.624	1.502
	(0.019)	(0.054)	(0.027)	(0.058)	(0.12)	(0.119)	(0.132)	(0.174)
	[1.189]	[2.838]	[1.357]	[2.805]	[2.157]	[2.022]		
Observations	3818	2740	2546	2354	322	288	3818	2740

Table 2
Effect of winning an election on subsequent party electoral success: alternative specifications, and refutability test, regression discontinuity estimates

Dependent variable	(1) Vote share $t+1$	(2) Vote share $t+1$	(3) Vote share $t+1$	(4) Vote share $t+1$	(5) Vote share $t+1$	(6) Res. vote share $t+1$	(7) 1st dif. vote share, $t+1$	(8) Vote share $t-1$
Victory, election t	0.077 (0.011)	0.078 (0.011)	0.077 (0.011)	0.077 (0.011)	0.078 (0.011)	0.081 (0.014)	0.079 (0.013)	−0.002 (0.011)
Dem. vote share, $t-1$	–	0.293 (0.017)	–	–	0.298 (0.017)	–	–	–
Dem. win, $t-1$	–	−0.017 (0.007)	–	–	−0.006 (0.007)	–	−0.175 (0.009)	0.240 (0.009)
Dem. political experience	–	–	−0.001 (0.001)	–	0.000 (0.003)	–	−0.002 (0.003)	0.002 (0.002)
Opp. political experience	–	–	0.001 (0.001)	–	0.000 (0.004)	–	−0.008 (0.004)	0.011 (0.003)
Dem. electoral experience	–	–	–	−0.001 (0.001)	−0.003 (0.003)	–	−0.003 (0.003)	0.000 (0.002)
Opp. electoral experience	–	–	–	0.001 (0.001)	0.003 (0.004)	–	0.011 (0.004)	−0.011 (0.003)

Maimonides' Rule

Angrist and Lavy look at the effects of school class size on kid's outcomes

Maimonides was a twelfth century Rabbinic scholar

He interpreted the Talmud in the following way:

> *Twenty-five children may be put it charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed.*

This rule has had a major impact on education in Israel

They try to follow this rule so that no class has more than 40 kids

But this means that

- If you have 80 kids in a grade, you have two classes with 40 each
- if you have 81 kids in a grade, you have three classes with 27 each

That sounds like a regression discontinuity

We can write the rule as

$$f_{sc} = \frac{e_s}{\left[int\left(\frac{e_s - 1}{40}\right) + 1\right]}$$

Ideally we could condition on grades with either 80 or 81 kids

More generally there are two ways to do this

- condition on people close to the cutoff and use f_{sc} as an instrument
- Control for class size in a “smooth” way and use f_{sc} as an instrument

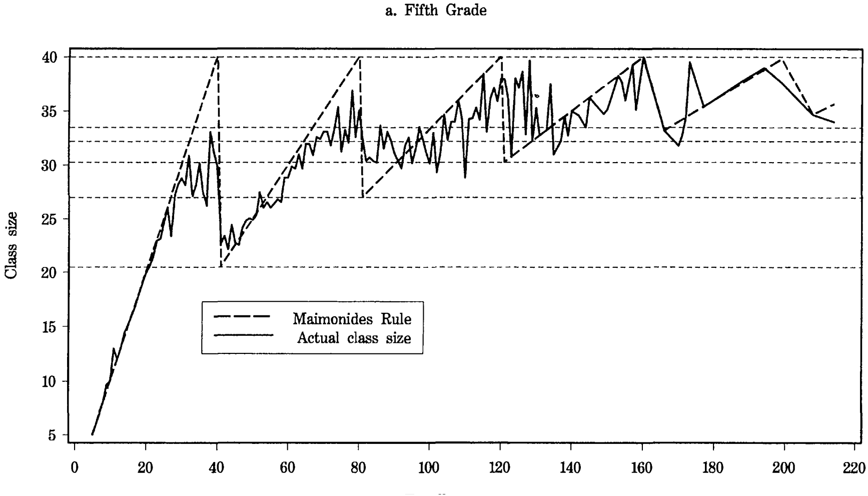
UNWEIGHTED DESCRIPTIVE STATISTICS

Variable	Mean	S.D.	Quantiles 0.10	0.25	0.50	0.75	0.90
A. Full sample							
5th grade (2019 classes, 1002 schools, tested in 1991)							
Class size	29.9	6.5	21	26	31	35	38
Enrollment	77.7	38.8	31	50	72	100	128
Percent disadvantaged	14.1	13.5	2	4	10	20	35
Reading size	27.3	6.6	19	23	28	32	36
Math size	27.7	6.6	19	23	28	33	36
Average verbal	74.4	7.7	64.2	69.9	75.4	79.8	83.3
Average math	67.3	9.6	54.8	61.1	67.8	74.1	79.4
4th grade (2049 classes, 1013 schools, tested in 1991)							
Class size	30.3	6.3	22	26	31	35	38
Enrollment	78.3	37.7	30	51	74	101	127
Percent disadvantaged	13.8	13.4	2	4	9	19	35
Reading size	27.7	6.5	19	24	28	32	36
Math size	28.1	6.5	19	24	29	33	36
Average verbal	72.5	8.0	62.1	67.7	73.3	78.2	82.0
Average math	68.9	8.8	57.5	63.6	69.3	75.0	79.4
3rd grade (2111 classes, 1011 schools, tested in 1992)							
Class size	30.5	6.2	22	26	31	35	38
Enrollment	79.6	37.3	34	52	74	104	129
Percent disadvantaged	13.8	13.4	2	4	9	19	35
Reading size	24.5	5.4	17	21	25	29	31
Math size	24.7	5.4	18	21	25	29	31
Average verbal	86.3	6.1	78.4	83.0	87.2	90.7	93.1
Average math	84.1	6.8	75.0	80.2	84.7	89.0	91.9

B. +/− 5 Discontinuity sample (enrollment 36–45, 76–85, 116–124)

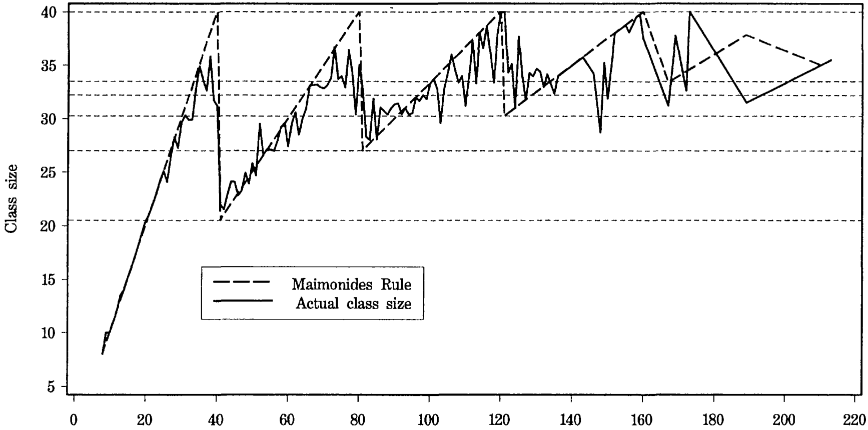
	5th grade Mean	5th grade S.D.	4th grade Mean	4th grade S.D.	3rd grade Mean	3rd grade S.D.
	(471 classes, 224 schools)		(415 classes, 195 schools		(441 classes, 206 schools)	
Class size	30.8	7.4	31.1	7.2	30.6	7.4
Enrollment	76.4	29.5	78.5	30.0	75.7	28.2
Percent disadvantaged	13.6	13.2	12.9	12.3	14.5	14.6
Reading size	28.1	7.3	28.3	7.7	24.6	6.2
Math size	28.5	7.4	28.7	7.7	24.8	6.3
Average verbal	74.5	8.2	72.5	7.8	86.2	6.3
Average math	67.0	10.2	68.7	9.1	84.2	7.0

Variable definitions are as follows: Class size = number of students in class in the spring, Enrollment = September grade enrollment, Percent disadvantaged = percent of students in the school from "disadvantaged backgrounds," Reading size = number of students who took the reading test, Math size = number of students who took the math test, Average verbal = average composite reading score in the class, Average math = average composite math score in the class.

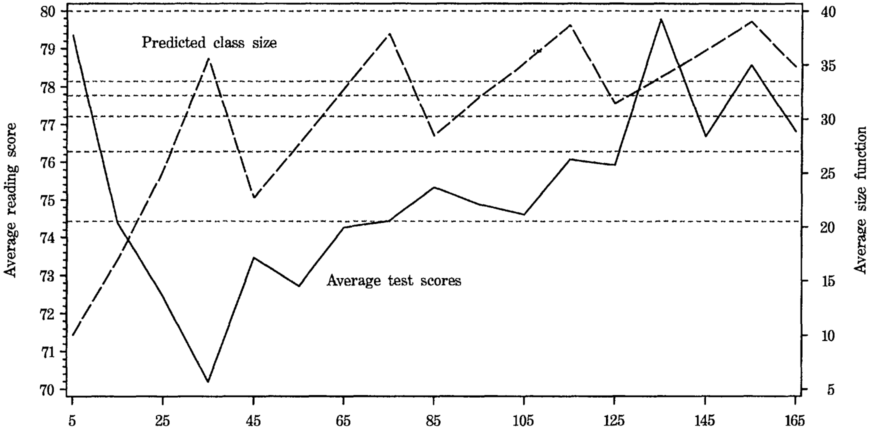


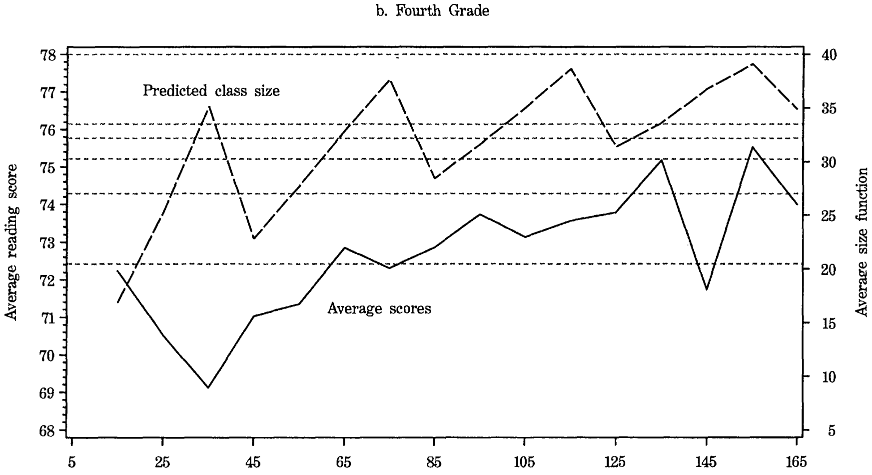
a. Fifth Grade
Class size
40
35
30
25
20
15
10
5
0
20
40
60
80
100
120
140
160
180
200
220
Maimonides Rule
Actual class size

b. Fourth Grade



a. Fifth Grade





b. Fourth Grade
Predicted class size
Average scores
Average reading score
Average size function
78
77
76
75
74
73
72
71
70
69
68
40
35
30
25
20
15
10
5
5
25
45
65
85
105
125
145
165

To estimate the model they use an econometric framework

$$Y_{ics} = \beta_0 + \beta_1 C_{cs} + \beta_2 X_{ics} + \alpha_s + \varepsilon_{ics}$$

Now we can't just put in a school effect because we will loose too much variation so think of α_s as part of the error term

Their data is a bit different because it is by class rather than by individual-but for this that isn't a big deal

Angrist and Lavy first estimate this model by OLS to show what we would get

TABLE II
OLS Estimates for 1991

	5th Grade						4th Grade					
	Reading comprehension			Math			Reading comprehension			Math		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score		74.3			67.3			72.5			69.9	
(s.d.)		(8.1)			(9.9)			(8.0)			(8.8)	
Regressors												
Class size	.221	−.031	−.025	.322	.076	.019	0.141	−.053	−.040	.221	.055	.009
	(.031)	(.026)	(.031)	(.039)	(.036)	(.044)	(.033)	(.028)	(.033)	(.036)	(.033)	(.039)
Percent disadvantaged		−.350	−.351		−.340	−.332		−.339	−.341		−.289	−.281
		(.012)	(.013)		(.018)	(.018)		(.013)	(.014)		(.016)	(.016)
Enrollment			−.002			.017			−.004			.014
			(.006)			(.009)			(.007)			(.008)
Root MSE	7.54	6.10	6.10	9.36	8.32	8.30	7.94	6.65	6.65	8.66	7.82	7.81
R^2	.036	.369	.369	.048	.249	.252	.013	.309	.309	.025	.204	.207
N		2,019			2,018			2,049			2,049	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes.

Next, they want to worry about the fact that C_{cs} is correlated with $\alpha_s + \varepsilon_{ics}$

They run instrumental variables using f_{sc} as an instrument.

TABLE IV
2SLS ESTIMATES FOR 1991 (FIFTH GRADERS)

	Reading comprehension						Math					
	Full sample				+/− 5 Discontinuity sample		Full sample				+/− 5 Discontinuity sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mean score	74.4				74.5		67.3				67.0	
(s.d.)	(7.7)				(8.2)		(9.6)				(10.2)	
Regressors												
Class size	−.158	−.275	−.260	−.186	−.410	−.582	−.013	−.230	−.261	−.202	−.185	−.443
	(.040)	(.066)	(.081)	(.104)	(.113)	(.181)	(.056)	(.092)	(.113)	(.131)	(.151)	(.236)
Percent disadvantaged	−.372	−.369	−.369		−.477	−.461	−.355	−.350	−.350		−.459	−.435
	(.014)	(.014)	(.013)		(.037)	(.037)	(.019)	(.019)	(.019)		(.049)	(.049)
Enrollment		.022	.012			.053		.041	.062			.079
		(.009)	(.026)			(.028)		(.012)	(.037)			(.036)
Enrollment squared/100			.005						−.010			
			(.011)						(.016)			
Piecewise linear trend				.136						.193		
				(.032)						(.040)		
Root MSE	6.15	6.23	6.22	7.71	6.79	7.15	8.34	8.40	8.42	9.49	8.79	9.10
N		2019		1961	471			2018		1960	471	

The unit of observation is the average score in the class. Standard errors are reported in parentheses. Standard errors were corrected for within-school correlation between classes. All estimates use f_{sc} as an instrument for class size.

Do Better Schools Matter? Parental Valuation of Elementary Education

Sandra Black, QJE, 1999

In the Tiebout model parents can "buy" better schools for their children by living in a neighborhood with better public schools

How do we measure the willingness to pay?

Just looking in a cross section is difficult: Richer parents probably live in nicer areas that are better for many reasons

Black uses the school border as a regression discontinuity

We could take two families who live on opposite side of the same street, but are zoned to go to different schools

The difference in their house price gives the willingness to pay for school quality.

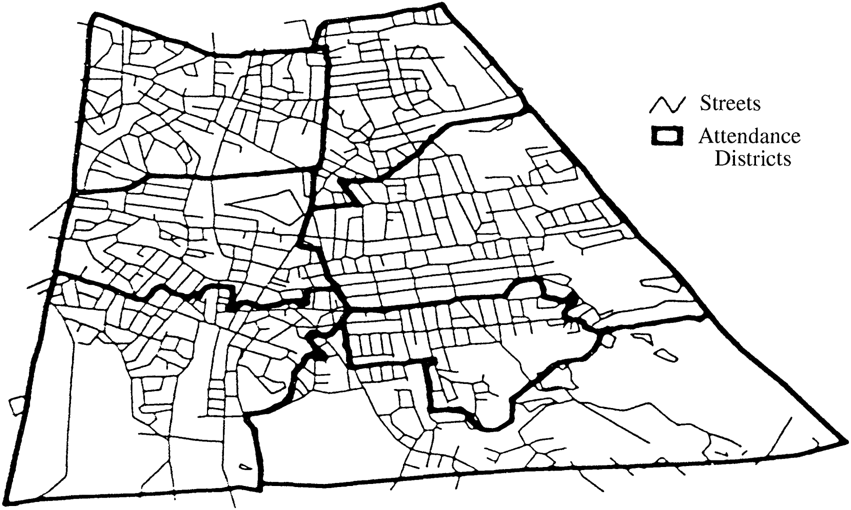


FIGURE I
Example of Data Collection for One City: Melrose Streets, and Attendance District Boundaries

TABLE II

REGRESSION RESULTS[a]

(ADJUSTED STANDARD ERRORS ARE IN PARENTHESES[b])

DEPENDENT VARIABLE = ln (HOUSE PRICE)

Distance from boundary:	(1) All houses[d]	(2) 0.35 mile from boundary (616 yards)	(3) 0.20 mile from boundary (350 yards)	(4) 0.15 mile from boundary (260 yards)	(5) 0.15 mile from boundary (260 yards)
Elementary school test score[c]	.035	.016	.013	.015	.031
	(.004)	(.007)	(.0065)	(.007)	(.006)
Bedrooms	.033	.038	.037	.033	.035
	(.004)	(.005)	(.006)	(.007)	(.007)
Bathrooms	.147	.143	.135	.167	.193
	(.014)	(.018)	(.024)	(.027)	(.028)
Bathrooms squared	−.013	−.017	−.015	−.024	−.025
	(.003)	(.004)	(.005)	(.006)	(.007)
Lot size (1000s)	.003	.005	.005	.005	.004
	(.0003)	(.0005)	(.0005)	(.0007)	(.0006)
Internal square footage (1000s)	.207	.193	.191	.195	.191
	(.007)	(.01)	(.01)	(.02)	(.012)
Age of building	−.002	−.002	−.003	−.003	−.002
	(.0003)	(.0002)	(.0005)	(.0006)	(.0004)
Age squared	.000003	.000003	.00001	.000009	.000005
	(.000001)	(.0000006)	(.000002)	(.000003)	(.000002)
Boundary fixed effects	NO	YES	YES	YES	NO
Census variables	Yes	No	No	No	Yes
N	22,679	10,657	6,824	4,594	4,589
Number of boundaries	N/A	175	174	172	N/A
Adjusted R^2	0.6417	0.6745	0.6719	0.6784	.6564

Differences in Means[a]

Distance from boundary:	Full sample		0.35 mile		0.20 mile		0.15 mile	
	Difference in means	*T*-statistic	Ratio of 0.35 to full sample[d]	*T*-statistic	Ratio of 0.20 to full sample[d]	*T*-statistic	Ratio of 0.15 to full sample[d]	*T*-statistic
ln (house price)	.045	3.82	0.85	3.32	0.85	3.17	0.93	3.17
Test score (sum of reading and math)	1.0	32.90	1.03	27.28	1.06	24.44	1.06	22.57
House characteristics								
Bedrooms	0.02	1.68	0.90	0.91	−0.35	−0.30	0.25	0.18
Bathrooms	0.03	2.98	0.23	0.52	−0.02	−0.05	−0.07	−0.12
Lot size	2011	11.39	0.22	2.14	0.24	1.95	0.12	0.83
Internal square footage	31	2.93	0.61	1.32	0.61	1.07	0.84	1.17
Age of building	−3.13	−6.92	0.75	−3.71	0.94	−3.76	1.09	−3.52
Neighborhood characteristics[c]								
Percent Hispanic	−.0008	−0.79	2.50	−1.35	2.50	−1.21	2.50	−1.26
Percent non-Hispanic black	−.0007	−1.50	0.43	−0.54	0.00	−0.07	−0.14	0.16
Percent 0–9 years old	.005	3.30	0.16	0.63	−0.08	−0.31	−0.30	−1.21
Percent 65+ years old	−.01	−2.04	0.40	−0.72	0.67	−1.28	0.60	−0.95
Percent female-headed households with children	−.001	−3.67	1.00	−3.17	1.20	−2.53	1.00	−2.38
Percent with bachelor's degree	.002	1.06	0.75	0.64	1.00	0.74	0.75	0.67
Percent with graduate degree	.008	3.32	0.88	2.77	0.88	3.02	0.88	3.31
Percent with less than high school diploma	−.005	−2.19	1.20	−2.02	0.80	−1.57	0.34	−0.64
Median household income	2,135	2.87	0.60	1.90	0.65	2.11	0.52	1.61

TABLE IV

MAGNITUDE OF RESULTS[a]

	(1) Basic hedonic regression[d]	(2) 0.35 sample boundary fixed effects	(3) 0.20 sample boundary fixed effects	(4) 0.15 sample boundary fixed effects
Coefficient on elementary school test score[b]	.035 (.004)	.016 (.007)	.013 (.0065)	.015 (.007)
Magnitude of effect (percent change in house price as a result of a 5% change in test scores)[c]	4.9%	2.3%	1.8%	2.1%
$ Value (at mean tax-adjusted house price of $188,000 in $1993)	$9212	$4324	$3384	$3948
$ Value (at median tax-adjusted house price of $158,000 in $1993)	$7742	$3634	$2844	$3318

a. The results presented here are based on estimates from Table II, columns (1)–(4).

b. Test scores are measured at the elementary school level and represent the sum of the reading and math scores from the fourth grade MEAP test averaged over three years (1988, 1990, and 1992). *Source*: Massachusetts Department of Education.

c. Approximately a one-standard-deviation change in the average test scores at the mean.

d. Regression includes house characteristics, school characteristics measured at the school district level, and neighborhood characteristics measured at the census block group level. See Table II, column (1), and Appendix 1 for more complete results.

Market Structure and Competition: Evidence from a Natural Experiment in Liquor Licensure

by Illanes and Moshary

This is a very recent paper that is still a working paper.

Many states in the U.S. regulate liquor stores where sales are done explicitly by the state

In 2012 Washington deregulated and allowed for private sale of liquor

They did this in a specific way though so that retailers were allowed to sell liquor as well as their premises exceeded 10,000 square feet

This leads to a natural regression discontinuity

They use this discontinuity to look at a number of different outcomes

Table 1: Summary Statistics for WSLCB Stores

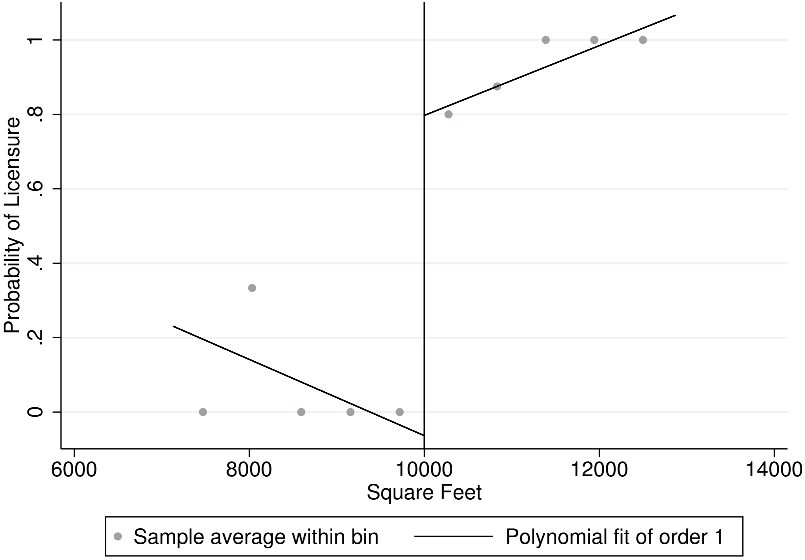
Summary Statistics for Beer, Wine and Liquor Licensure	
Prior to 2012: Beer and wine licensed retailers	4,978
Chain licensees	2,098
At Liberalization: Existing Beer/Wine Licensees	4,977
Liquor-licensed	1,075
Chain liquor licensees	924
At Liberalization: Entrants	570
Liquor-licensed	57
Beer and wine licensed	558
Chain stores	130

Entry

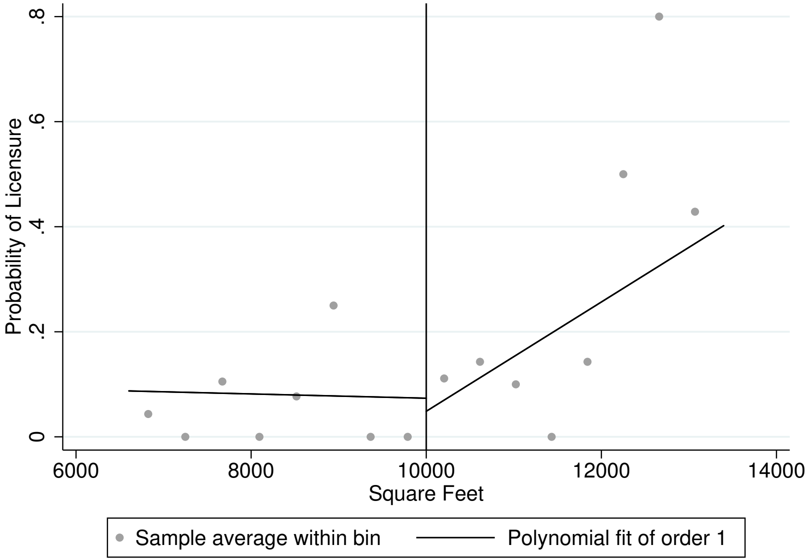
The first thing to see is whether there was actually an effect of the deregulation on entry into the market.

They use the specification (their notation)

$$
\begin{aligned}
1\left[\text{Liquor Licensed}\right]_s = & \alpha_0 + \alpha_1 1\left[SqFt_s \geq 10000\right]_s + \alpha_2 SqFt_s \\
& + \alpha_3 1\left[SqFt_s \geq 10000\right]_s SqFt_s + \varepsilon_s
\end{aligned}
$$



(a) Chain Stores



(b) Independent Stores

Table 2: Regression Discontinuity Estimates of the Effect of License Eligibility on Entry

RD Estimates of the Effect of Licensure on Entry				
	(1)	(2)	(3)	(4)
	All Stores	Independent Stores	Chain Stores	Large Chains (10+ Stores)
Licensure Discontinuity	0.26**	-0.03	0.86***	0.88***
	(0.112)	(0.133)	(0.153)	(0.160)
Observations	4605	2599	2006	1870
Effective Observations – Below	194	102	103	23
Effective Observations – Above	130	87	55	40
Bandwidth	4149.9	3634.8	3397.6	2867.5
McCrary Test P-Value	0.379	0.620	0.545	0.981

Entry of competitors

The next question is about competition

How does my entry depend on potential competitors

Here we use the regression discontinuity not for my store size but rather competitors near by

$$
\begin{aligned}
Y_s = & \alpha_0 + \alpha_1 1\left[\text{IsChain}\right]_s + \alpha_2 N_s^{d,10-15} + \\
& \alpha_3 1\left[\text{IsChain}\right]_s N_s^{d,10-15} + \sum_k \lambda_k^d 1\left[N_s^{d,5-15} = k\right] + \varepsilon_s
\end{aligned}
$$

Table 3: Effect of License Eligibility of Nearby Stores on Own Entry Decisions

		Effect of the License Eligibility of Nearby Stores on Own Entry Decisions									
		Bandwidth = 5000 square feet									
stance to Store (miles):		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Independents	# of Marginally License Eligible Neighbors	-0.158	-0.218***	-0.181***	-0.170***	-0.114**	-0.102***	-0.064*	-0.067*	-0.027	-0.045
		(0.107)	(0.068)	(0.058)	(0.044)	(0.046)	(0.035)	(0.038)	(0.034)	(0.033)	(0.029)
	Baseline Entry Probability	0.323***	0.340***	0.345***	0.354***	0.349***	0.354***	0.346***	0.351***	0.341***	0.354***
		(0.025)	(0.025)	(0.026)	(0.027)	(0.027)	(0.027)	(0.028)	(0.029)	(0.030)	(0.031)
Chains	# of Marginally License Eligible Neighbors	0.073	0.012	-0.002	-0.009	-0.007	-0.001	0.002	0.014	0.012	0.002
		(0.051)	(0.036)	(0.036)	(0.032)	(0.027)	(0.021)	(0.021)	(0.019)	(0.017)	(0.016)
	Baseline Entry Probability	0.948***	0.951***	0.953***	0.954***	0.954***	0.952***	0.951***	0.947***	0.945***	0.950***
		(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.012)
# of Neighbors in the Bandwidth FE		x	x	x	x	x	x	x	x	x	x
N		1173	1173	1173	1173	1173	1173	1173	1173	1173	1173

Table 4: Effect of License Eligibility of Nearby Stores on Own Liquor Revenue

		Effect of the License Eligibility of Nearby Stores on Own Sales of Liquor Bandwidth = 5000 square feet									
istance to Store (miles):		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Independents	# of Marginally License Eligible Neighbors	74,164***	26,344	9,939	23,061	17,613	15,317	30,489*	15,789	27,740**	12,942
		(26,994)	(19,522)	(17,099)	(19,248)	(21,375)	(14,965)	(15,994)	(13,241)	(12,096)	(10,543)
	Baseline Sales	28,600***	29,876***	30,960***	29,757***	29,293***	30,063***	24,697***	28,024***	23,001***	27,330***
		(4,616)	(5,086)	(5,407)	(5,148)	(5,402)	(5,129)	(5,773)	(5,607)	(5,815)	(6,161)
Chains	# of Marginally License Eligible Neighbors	125,407***	60,278**	52,321**	48,610**	65,412***	48,615**	45,496***	36,894**	40,954***	35,933***
		(33,004)	(26,185)	(24,356)	(24,185)	(22,696)	(19,144)	(17,682)	(14,994)	(13,958)	(13,100)
	Baseline Sales	245,564***	246,837***	245,818***	243,644***	238,400***	238,942***	236,633***	237,725***	232,210***	233,325***
		(9,374)	(9,778)	(10,184)	(10,434)	(10,085)	(10,302)	(10,476)	(10,786)	(9,714)	(9,981)
# of Neighbors in the Bandwidth FE		x	x	x	x	x	x	x	x	x	x
N		1173	1173	1173	1173	1173	1173	1173	1173	1173	1173

Consumer Outcomes

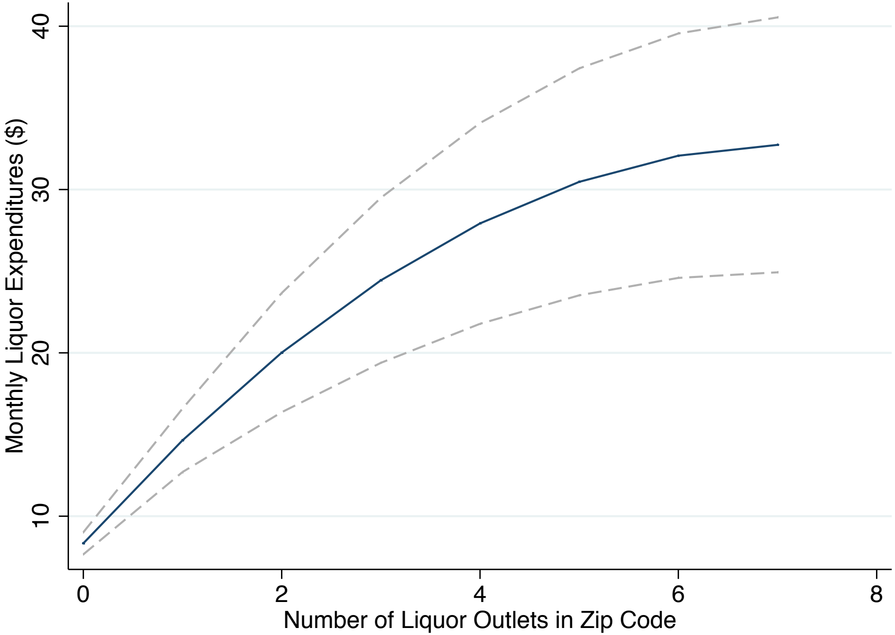
We can look at the number of stores on outcomes for consumers

$$y_{ht} = \alpha_0 + \alpha_1 NL_{z(h,t)} + \alpha_2 NL^2_{z(h,t)} + X'_{z(h,t)}\delta + \varepsilon_{ht}$$

Here they use the RD as an instrument. That is they use control for the number of stores near by between 5000 and 15000 square feet using the number of size bigger than 10000 as the instrument (and interact this with the number of large stores for the squared term)

Table 5: Effect of License Eligibility on Purchasing

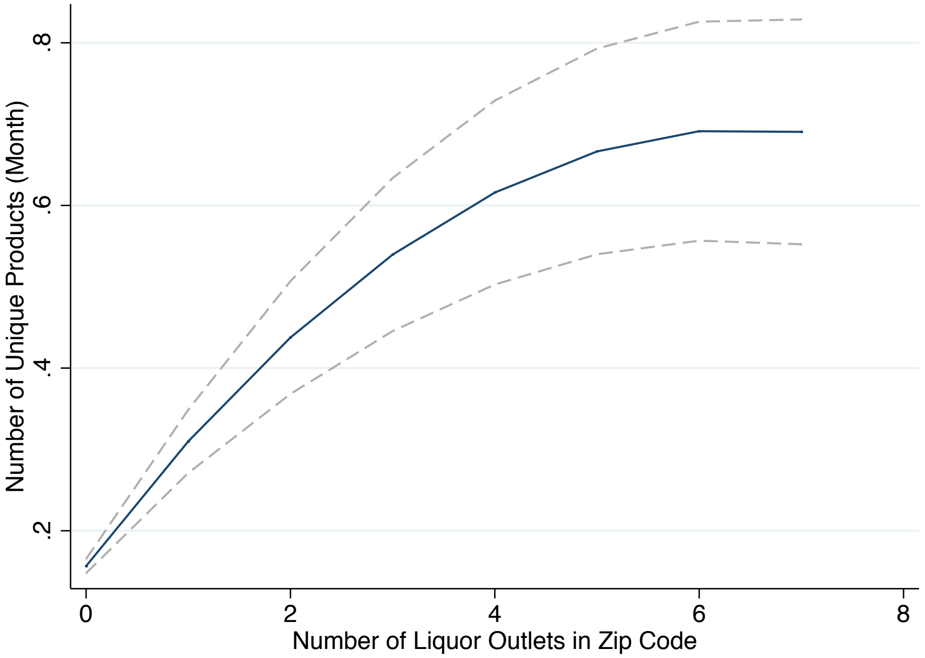
Effect of Market Structure on Consumption			
	(1)	(2)	(3)
Panel A: IV	Expenditures ($)	Volume (L)	Ethanol (L)
# of Liquor Retailers	6.214**	0.215**	0.089**
	(2.637)	(0.090)	(0.036)
# of Liquor Retailers^2	-0.418**	-0.015**	-0.006**
	(0.191)	(0.007)	(0.003)
# of Stores in the Bandwidth FE	X	X	X
# of Stores Above the Bandwidth FE	X	X	X
Mean	7.875	0.271	0.108
Observations	31875	31875	31875
Panel B: Reduced Form	Expenditures ($)	Volume (L)	Ethanol (L)
# of Marginally License-Eligible Stores	6.231**	0.204**	0.086**
	(2.841)	(0.100)	(0.039)
# of Marginally License-Eligible Stores × # Stores Above the Bandwidth	-0.704***	-0.024***	-0.010***
	(0.249)	(0.009)	(0.004)
# of Stores in the Bandwidth FE	X	X	X
# of Stores Above the Bandwidth FE	X	X	X
Observations	31875	31875	31875



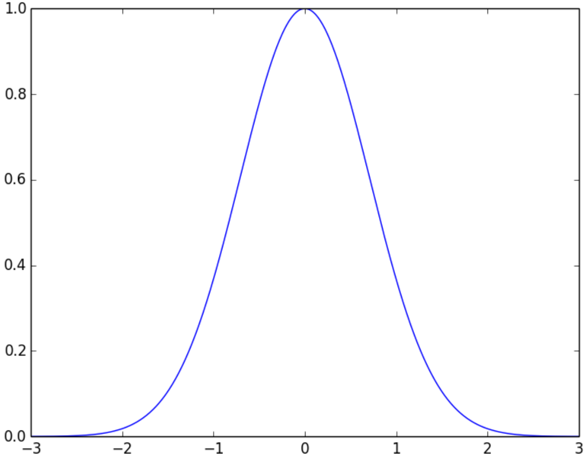
Monthly Liquor Expenditures ($)
10
20
30
40
0
2
4
6
8
Number of Liquor Outlets in Zip Code

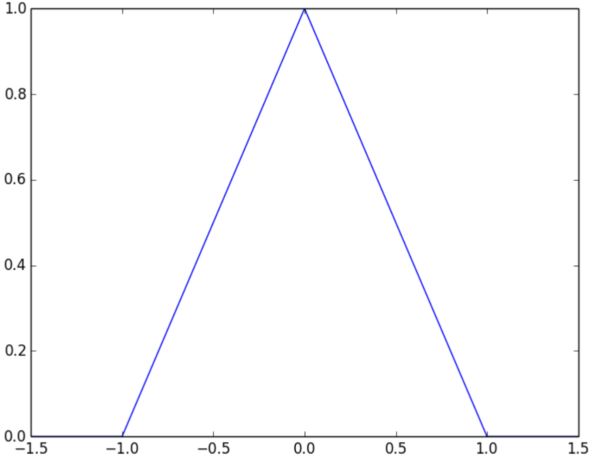
Table 7: Effect of Market Configuration on Prices

Effect of Market Structure on Log Price				
Panel A: IV	(1)	(2)	(3)	(4)
# of Liquor Outlets	0.079	-0.000	0.042	-0.018
	(0.083)	(0.013)	(0.113)	(0.017)
# of Liquor Outlets^2			-0.003	0.002
			(0.008)	(0.001)
# of Stores in the Bandwidth FE	X	X	X	X
# of Stores above the Bandwidth FE	X	X	X	X
UPC FE		X		X
Panel B: Reduced Form				
# of Marginally License-Eligible Stores	0.056	-0.000	0.040	-0.007
	(0.052)	(0.012)	(0.059)	(0.016)
# of Marginally License-Eligible Stores × # Stores above the Bandwidth			0.003	0.001
			(0.007)	(0.002)
# of Stores in the Bandwidth FE	X	X	X	X
# of Stores above the Bandwidth FE	X	X	X	X
UPC FE		X		X
Observations	6027	6027	6027	6027

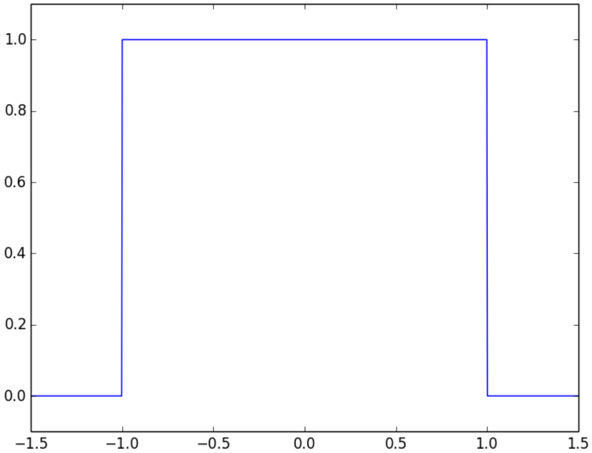


Number of Unique Products (Month)
.2
.4
.6
.8
0
2
4
6
8
Number of Liquor Outlets in Zip Code

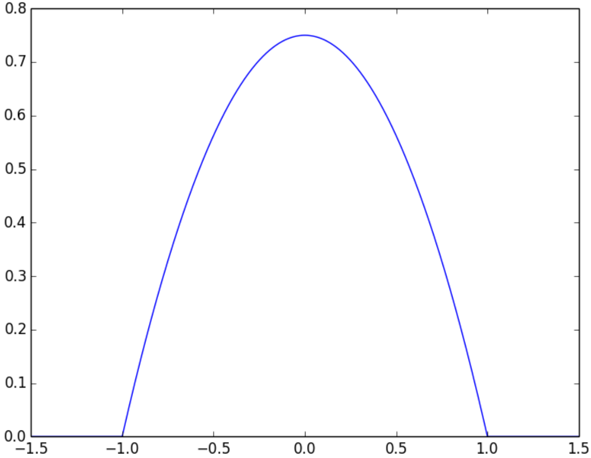

1.0
0.8
0.6
0.4
0.2
0.0
−3
−2
−1
0
1
2
3



Back



Back


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5