Empirical Applications
Regression Discontinuity

Part 1: The Electoral Advantage of Incumbency

1. The files electoral1.dta and electoral2.dta contain data to replicate the results in Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. Journal of Econometrics, 142(2), 675-697. The file electoral1.dta includes observations from 1946 to 1998 on individual candidates running for congressional office. The file electoral2.dta includes observations from 1946 to1998 on the outcomes of congressional district races. You can find variable definitions using the describe command in STATA. Read the article by Lee (2008) prior to conducting the analysis.

1.1. What is the key causal question that Lee (2008) attempting to answer? Why is the answer to this question interesting from a policy perspective?

The key causal question deals with whether being an incumbent in a given congressional district directly increases the chance of winning an election. This is interesting from a policy perspective because many people claim that incumbents use their position unjustly to win elections while others claim that incumbents have a higher chance because they are better than their competitors to begin with .The descriptive statistics shows that incumbents have a higher chance of winning an election. However, it is hard to draw a causal inference simply from the descriptive numbers. If we could identify a positive causal link between being an incumbent and winning elections, then we would provide some evidence to show that incumbents are winning elections because they are inherently better than their competitors and not because they are unfairly abusing their privileged position.

1.2. Explain why the conditional independence assumption is unlikely to hold in this case. Specifically, why might we be concerned that simple regression estimates of incumbency advantage might suffer from selection bias? What is the likely direction of this bias? Discuss this in terms of equation (3) of Lee (2008).

$$
\begin{aligned}
&v_{i2} = \alpha w_{i1} + \beta v_{i1} + \gamma d_{i2} + e_{i2},\\
&d_{i2} = 1\left[v_{i1} \geqslant \frac{1}{2}\right],\\
&f_{i1}(v|w) \text{ —density of } v_{i1} \text{ conditional on } w_{i1} \text{ —is continuous in } v,\\
&\mathrm{E}[e_{i2}|w_{i1}, v_{i1}] = 0,
\end{aligned}
$$

The Conditional Independence Assumption (CIA) states that, conditional on a set of covariates, the treatment is as good as randomly assigned. In our case, however, treatment is not as good as randomly assigned as the incumbent is more likely to win the next election. Lee (2008) argues that looking at a narrow range of votes shares of closely contested elections, the CIA could be fulfilled.

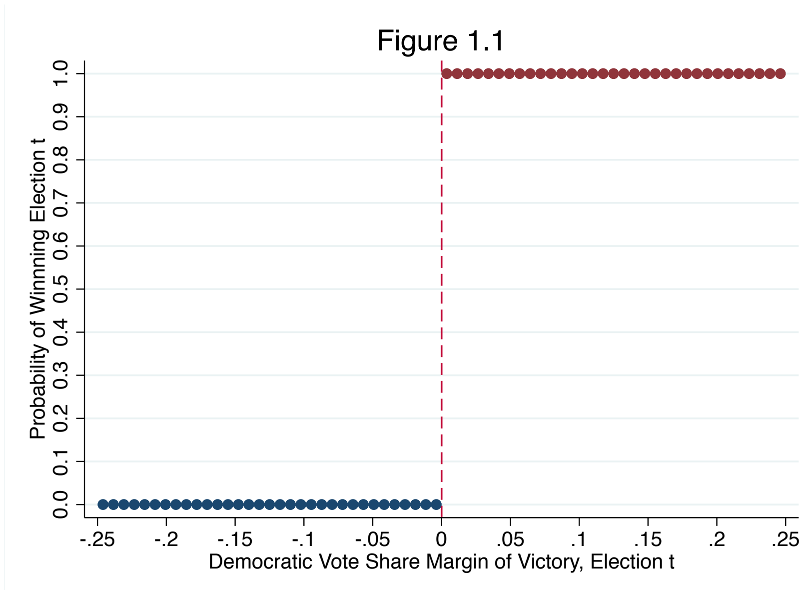
The very first line in the above equations from Lee's paper show a simple OLS regression model which determines the incumbency advantage for winners in the previous election cycle. The term v_{i2} represents vote share in the next election period and d_{i2} is an indicator for winning in next election cycle. The term w_{i1} represents a set of characteristics an incumbent might have such as more campaign financing and v_{i1} represents the vote share of the the winner in the previous election cycle. These two terms are correlated and since it is impossible to account for all the variables that would go into w_{i1}, we will have an omitted variable bias. Thus OLS regression will not lead us to make causal inference. Selection bias is an issue in this case because the winners in the election cycle t are fundamentally different from their competitors in election cycle t+1. The coefficient on d_{i2} will overestimate or be biased upwards due to omitted variables in the error term (Lee, 2008, p. 684).

1.3. Lee (2008) attempt to derive causal estimates of the effect of incumbency on future election outcomes using a regression discontinuity design. Briefly discuss their RD design and the conditions that must be met for the RD to provide unbiased estimates?

The Regression Discontinuity Design (RDD) in Lee's paper is that there is a random assignment of treatment (winning an election) when there is a very close contest between a democrat and an opponent. If the election is closely contested, a regression discontinuity could be employed by using a small interval of the vote shares by centering the running variable at democratic vote share minus the republican opponent votes share. Lee (2008) does not use a simple 50% cutoff because a candidate can win with a less percentage if there are more than two candidates.

For an RDD to work, there are four assumptions. The first assumption is that people cannot manipulate the situation to be on either side of the cutoff around the cutoff. In our case, candidates certainly plan to be on the winning side. However, Lee(2008) argues, in a very closely contested election, there is an element of randomness to winning which is amenable to an RDD analysis. The second assumption is that treatment is determined independent of the running variable. In our case, winning is determined by the vote percentage received by the democrat candidate. If the democrat candidate has the highest vote share, there is treatment . There is no treatment in all other cases. This fulfills the second assumption as winning an election is determined by a fixed rule that cannot be changed after the fact. The third assumption is that there is only a single factor (the treatment) that changes across the cutoff point. In our case, the only thing that changes is the vote share and thus the third assumption is met. The fourth assumption is that the function that relates the dependent variable to the running variable has a distribution that is continuous. This assumption is also fulfilled in our case assuming parties do not engage in fraudulent activities as the vote share has a continuous density conditional on covariates.

1.4. Using the data in electoral1.dta, create a graph showing the probability of winning an election in period t (win) as a function of the Democratic vote share margin of victory, election t (mov). Use STATAs cmogram command to create the graph. As in Lee (2008) Figure 2, restrict the x axis to the range -0.25 to 0.25 and label the x axis in increments of 0.05. Set the range of the y axis from 0 to 1 and label the y axis in increments of 0.10. Include a scatter line at the discontinuity (at zero) and use the cut(0) option. Is this a sharp or fuzzy RD design, explain.



This is a sharp RD design. This is to be expected as the candidate with the largest share of votes in the current period always get the elected position (wins) in the current period.

1.5. Replicate Figures 2a, 2b, 3a, 3b in Lee (2008). Use the qfit option in cmogram to add quadratic trends to the figures. Make sure you adjust the units on your axis in these diagrams to match those in Lee (2008). Discuss the implications of these graphs. For example, what does Figure 2a suggest?

As can be seen on the following page, Figures 1.2-1.5 suggest that there is a discontinuity in the probability of winning and probability of candidacy in the next election. Figure 2a (Figure 1.2) specifically suggests that the probability of winning in the next election t+1 is much higher for candidates who won election t.

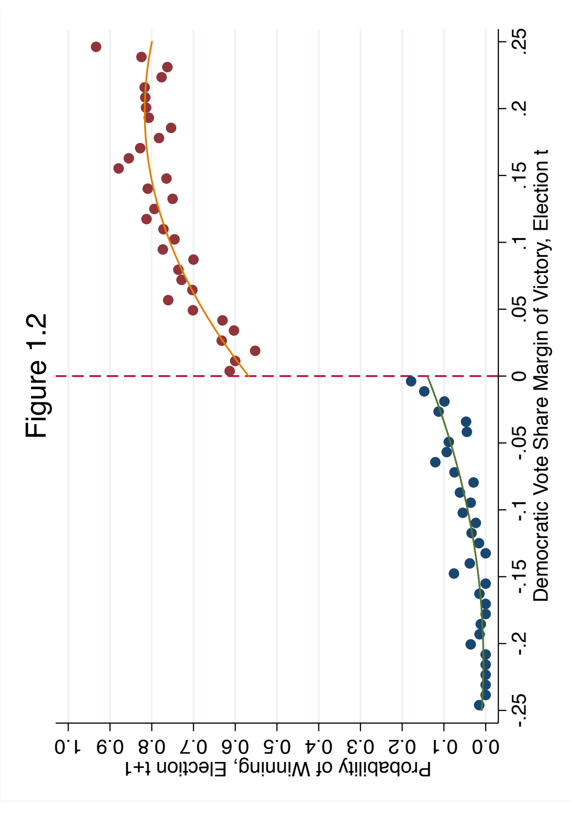


Figure 1.2

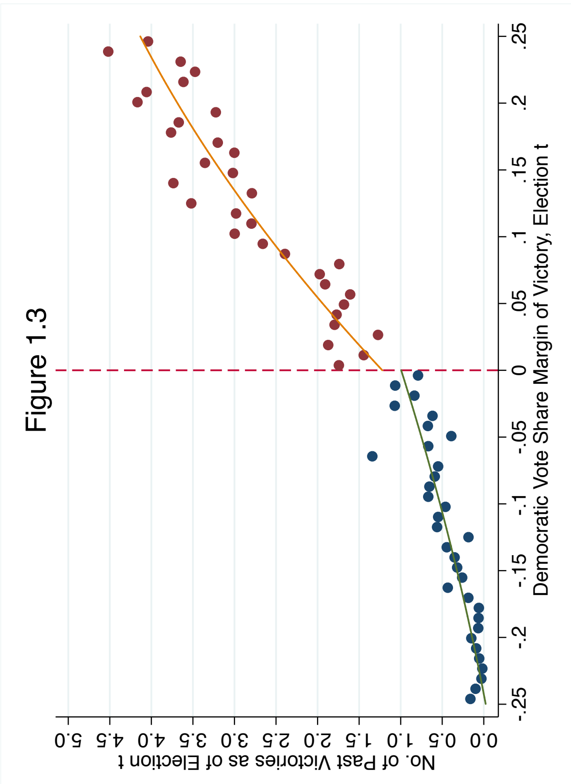


Figure 1.3

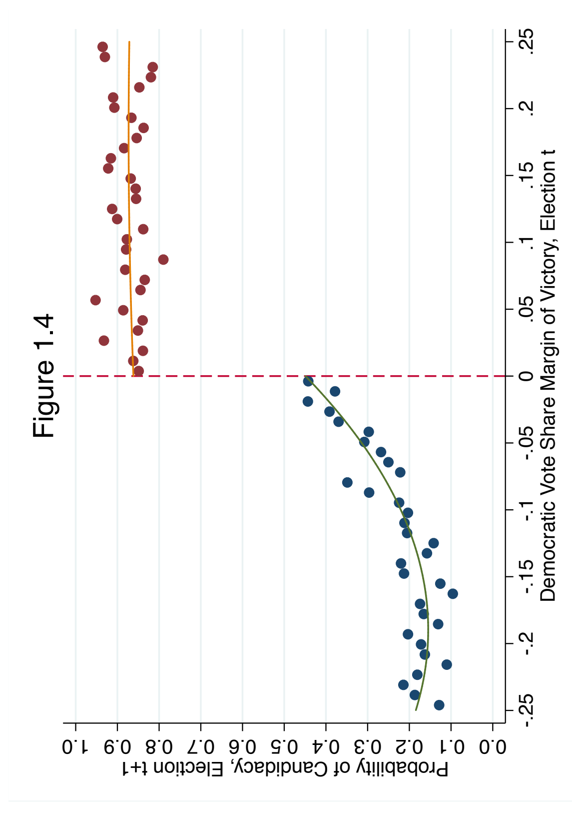


Figure 1.4

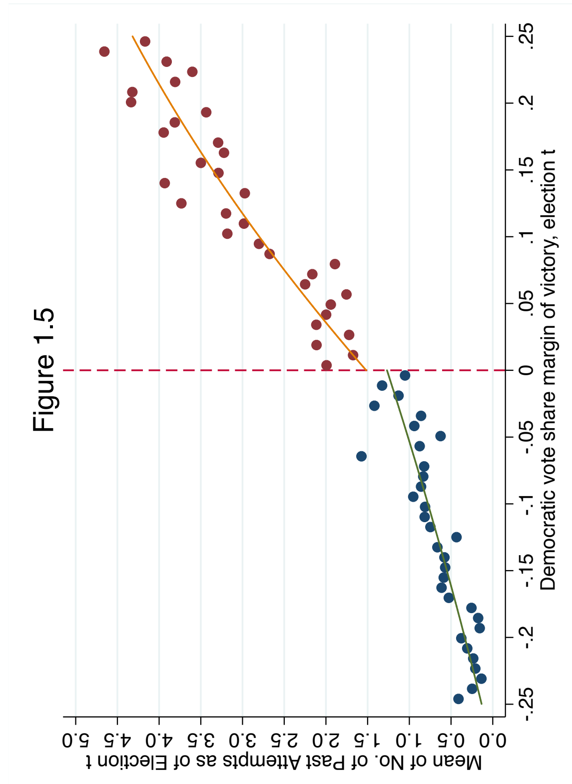


Figure 1.5

1.6. From Figure 2a we saw that barely winning an election in period t leads to a large increase in the probability of winning again in period t+1. Now we want to estimate that effect directly and get standard errors. Start by creating 2nd, 3rd and 4th order polynomials in the forcing variable and then also create interactions between those variables and the win indicator variable to allow for different trends on either side of the discontinuity.

Please see attached do file.

1.7. Now estimate models where the dependent variable is demsharenext and the independent variable of interest is the win indicator. Create a table showing the estimated coefficient on the win indicator from regressions that sequentially add higher order polynomials of the mov variable to the regression (i.e. start with only including the mov variable and then adding mov squared, etc). Finally also show the estimated coefficient on the win indicator from a regression that includes all the higher order terms and their interactions with the win indicator. Discuss the results and interpret the estimated coefficient on the win indicator. Which estimates to you find most credible? How do they compare to Figure 2a in Lee (2008)?

As can be seen in Table 1.1 below, using win and mov as the only regressors, we observe that the probability of winning an election by a small margin in election period t results in a 67.7% higher probability of winning election t+1. Using win, mov and squared move as the regressors, we observe that the probability of winning an election by a small margin in election period t results in a 65.9% higher probability of winning election t+1. Using win, mov squared, and move cubed as the regressors, we observe that the probability of winning an election by a small margin in election period t results in a 60.9% higher probability of winning election t+1.Using win, mov squared, move cubed, and move raised to the fourth level as the regressors, we observe that the probability of winning an election by a small margin in election period t results in a 59.8% higher probability of winning election t+1. Finally, when we add all the interactions between win and the polynomials of the mov variable to the last regression, we observe that the probability of winning an election by a small margin in election period t results in a 44.9 higher probability of winning election t+1.

We find the results in the last column to be most convincing as the regression best fits the underlying relationship between the outcome and the running variable. The result in the last column shows .449 causal effect which is almost the same as the discontinuity jump that we observed in Figure 1.2.

Table 1.1: Estimates of Various Models on Incumbency Advantage Indicator

VARIABLES	(1) Democtratic Vote Share of Margin, election t+1	(2) Democtratic Vote Share of Margin, election t+1	(3) Democtratic Vote Share of Margin, election t+1	(4) Democtratic Vote Share of Margin, election t+1	(5) Democtratic Vote Share of Margin, election t+1
Victory, election t	0.677***	0.659***	0.609***	0.598***	0.449***
	(0.00321)	(0.00295)	(0.00330)	(0.00575)	(0.00476)
Democratic vote share margin of victory, election t	0.159***	0.204***	0.363***	0.407***	1.381***
	(0.00539)	(0.00476)	(0.00733)	(0.0212)	(0.0423)
Squared Democratic vote share margin of victory, election t		-0.0844***	0.0955***	0.0179	4.790***
		(0.00654)	(0.00882)	(0.0187)	(0.199)
Cubed democratic vote share margin of victory, election t			-0.406***	-0.553***	6.800***
			(0.0163)	(0.0757)	(0.364)
Democratic vote share margin of victory, election t, raised to 4th				0.204***	3.297***
				(0.0780)	(0.225)
Interaction between Democratic vote share margin of victory, election t and victory, election t					0.659***
					(0.0743)
Interaction between squared Democratic vote share margin of victory, election t and victory, election t					-10.69***
					(0.351)
Interaction between cubed Democratic vote share margin of victory, election t and victory, election t					0.249
					(0.616)
Interaction between Democratic vote share margin of victory, election t (mov) raised to the 4th and victory, election t					-6.278***
					(0.350)
Constant	0.0628***	0.0803***	0.0908***	0.1000***	0.141***
	(0.00143)	(0.00145)	(0.00134)	(0.00307)	(0.00283)
Observations	9,674	9,674	9,674	9,674	9,674
R-squared	0.967	0.968	0.973	0.973	0.978

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

1.8. Now restrict the bandwidth to areas around the discontinuity and estimate models where the dependent variable is once again demsharenext and the independent variables include only the win indicator and a quadratic in the forcing variable. Use bandwidths of (-0.2, 0.2), (-0.1, 0.1), (-0.05, 0.05), (-0.02, 0.02). Create a table of your results. Comment on the difference between the estimates. In particular, what happens to the estimated coefficient on the win indicator as you narrow the bandwidth and what happens to the standard errors? Explain the changes.

As can be seen in Table 1.2 below, when we move from a bandwidth of (-0.2,0.2) to a bandwidth of (-0.02, 0.02), the coefficient on the win indicator decreases from a probability of 67.9% to 44.4%. This is because we are looking at a narrower range around the cutoff point for discontinuity. When we use a higher bandwidth, we include data point further out from the cutoff point which overestimates the coefficient on the win indicator. Using a narrower bandwidth gives us a more accurate estimate of the coefficient on the win indicator. The standard errors show an increasing trend overall as we move from a broader bandwidth to a narrower bandwidth. This demonstrates the tradeoff that we have to make between bias and uncertainty.

Table 1.2: Estimates on the Incumbency Advantage Indicator with Varying Bandwidths

	Bandwidth (.2)	**Bandwidth (.1)**	**Bandwidth (.05)**	**Bandwidth (.02)**
VARIABLES	Democratic Vote Share of Margin, election t+1	Democratic Vote Share of Margin, election t+1	Democratic Vote Share of Margin, election t+1	Democratic Vote Share of Margin, election t+1
Victory, election t	0.679***	0.588***	0.524***	0.444***
	(0.00246)	(0.00319)	(0.00446)	(0.00390)
Squared Democratic vote share margin of victory, election t	1.137***	4.539***	3.516	-228.1***
	(0.110)	(0.587)	(3.040)	(14.43)
Constant	0.0376***	0.0694***	0.0964***	0.186***
	(0.00213)	(0.00290)	(0.00409)	(0.00330)
Observations	3,743	1,905	947	354
R-squared	0.954	0.947	0.935	0.972

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

1.9. Now use STATAs rd command to estimate local linear regressions where the dependent variable is once again demsharenext. Use a bandwidth of 0.05. Repeat using STATAs rdrobust command with no bandwidth specified. Compare the results to your previous results.

Both the rd and rdrobust commands provide coefficients that are around the 40 percent range which is close, although lower, to the results we obtained in the last columns of the previous two tables. This shows that the estimates on incumbency advantage are similar when using both parametric and nonparametric (local linear regression) RDD analysis methods.

Table 1.3: RD Estimate						
Democtratic Vote Share of Margin, election t+1	Coef.	Std. Err.	z	P>z	[95% Conf	Interval]
lwald	0.3990111	0.0062948	63.39	0	0.386674	0.411349
lwald50	0.3888139	0.0074204	52.4	0	0.37427	0.403358
lwald200	0.4268928	0.0057661	74.03	0	0.415591	0.438194

rdrobust estimate						
Cutoff c = 0	Left of c	Right of c	Number of obs		=	9674
			BW type		=	mserd
Number of obs	4342	5332	Kernel		=	Triangular
Eff. Number of obs	571	580	VCE method		=	NN
Order loc. poly. (p)	1	1				
Order bias (q)	2	2				
BW loc. poly. (h)	0.06	0.06				
BW bias (b)	0.167	0.167				
rho (h/b)	0.361	0.361				
Outcome:	**Democrat Vote share of Margin, t+1**		**Running variable:**	**Democrat Vote Share of Margin, election t**		
Method	Coef.	Std. Err.	z	P>z	[95% Conf	Interval]
Conventional	0.40299	0.00109	368.6978	0	0.400846	0.405131
Robust	-	-	347.3479	0	0.395918	0.400412

1.10. Now let's turn to the data in electoral2.dta where the unit of observation is the Congressional district. Our goal now is to estimate the "incumbency advantage" which Lee (2008) defines as the overall causal impact of being the current incumbent party in a district on the votes obtained in the district's election. Lee also notes that estimation of the analogous effect for the individual candidate (i.e. data in electoral1.dta) is complicated by selective "drop-out". Briefly describe what Lee (2008) means by drop-out and why it impedes the estimation of causal effects.

Dropout refers to a situation whereby candidates do not run again in the next election period in cases where they either won or lost in the previous period. This happens because there are no rules that force candidates to run again. From a causal analysis perspective, this creates issues with “selective attrition” that makes drawing causal conclusions very difficult even in a well designed randomized experiment. Thus, Lee uses district level data that compares parties (Democrats and Republicans) instead of individual candidates (Lee, 2008, p. 685).

1.11. Start by replicating Figures 4a, 4b, 5a, 6b in Lee (2008). Use the qfit option in cmogram to add quadratic trends to the figures. Make sure you adjust the units on your axis in these diagrams to match those in Lee (2008). Interpret the graphs and their implications.

As we can see on the following page, Figure 1.6 shows there is a positive relationship between vote share in election period t+1 and vote share of margin in election t. Also the graph shows a sharp discontinuity at cutoff point which suggest a strong “incumbency advantage.” Figure 1.7 also show a strong positive relationship between probability of victory in election period t+1 and vote share of margin in election t. In addition, the graph shows a sharp discontinuity at cutoff point which suggest a strong “incumbency advantage.” Figures 1.8 and 1.9 show no discontinuity at the cutoff point. This is to be expected as the outcome variables in figures 1.7 and 1.9 are vote share, election t-1 and probability of victory, election t-1 respectively, which are covariates which should not be affected by vote share of margin in election t.

1.12. Create an indicator for whether the Democratic party wins in election t, quadratics in the running variable and interactions between the running variables and the win indicator. Replicate Table 2 in Lee (2008) and discuss the results.

The first five columns in Table 1.4 (p. 11) show the estimates on incumbency advantage when regressed on different combinations of covariates. The coefficient on the win indicator is of particular interest and measures the impact of winning election t on vote share in election t+1. As can be seen in the first row of the table, the coefficient on the win indicator stays around the 8% range. Using different combinations of regressors does not seem to impact the coefficient. In column 6, we see that the residuals of the regression in column 5 regressed against covariates from election period in t-1. Again, the coefficient is around 7.4% which demonstrates that the “...treatment is locally independent of all predetermined characteristics.” Column 7 has the difference between vote share, election t+1 and vote share, election t-1 as its dependent variable and the coefficient on the win indicator is around 8%. Column 8 has vote share, election t-1 as its dependent variable and the coefficient on the win indicator is very small and statistically insignificant. This is to be expected as vote share, election t should not have an impact vote shares in election t-1 (Lee, 2008, p. 690).

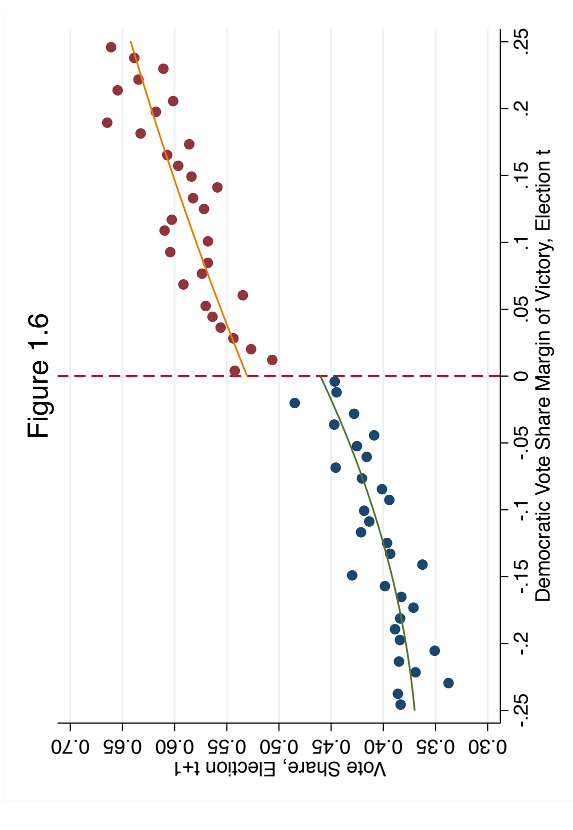


Figure 1.6

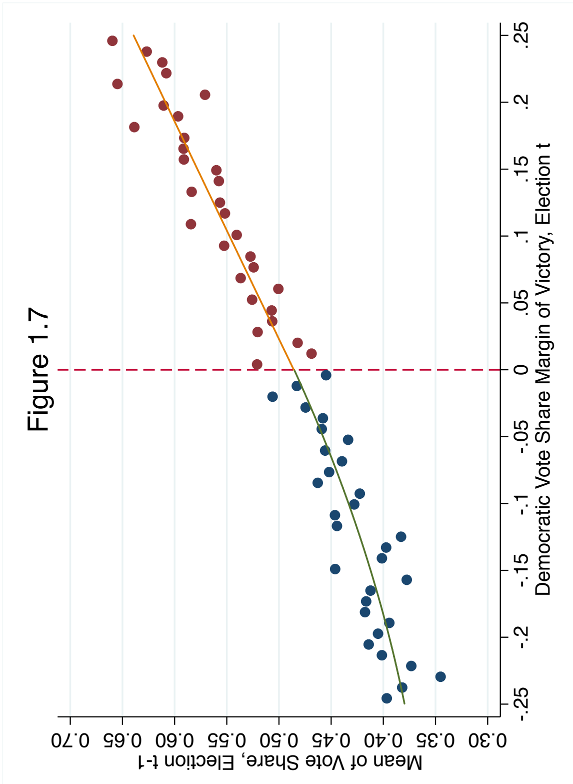


Figure 1.7

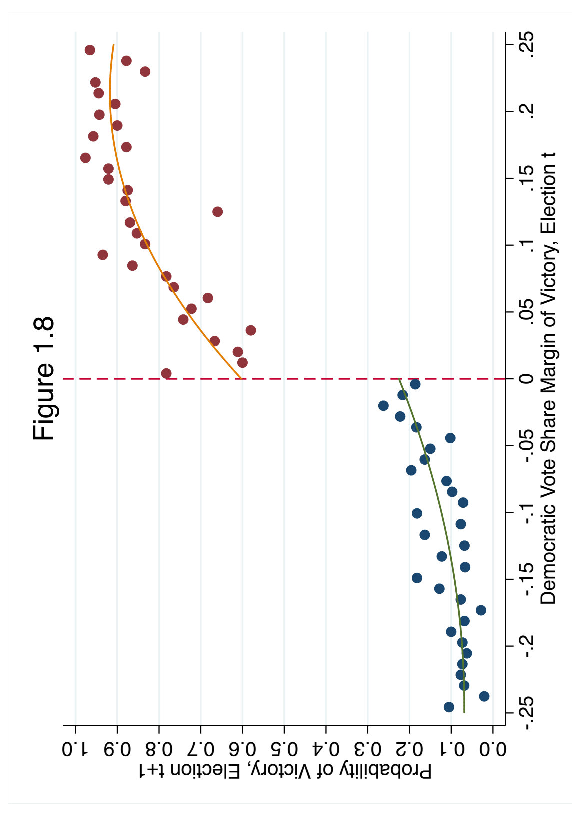


Figure 1.8

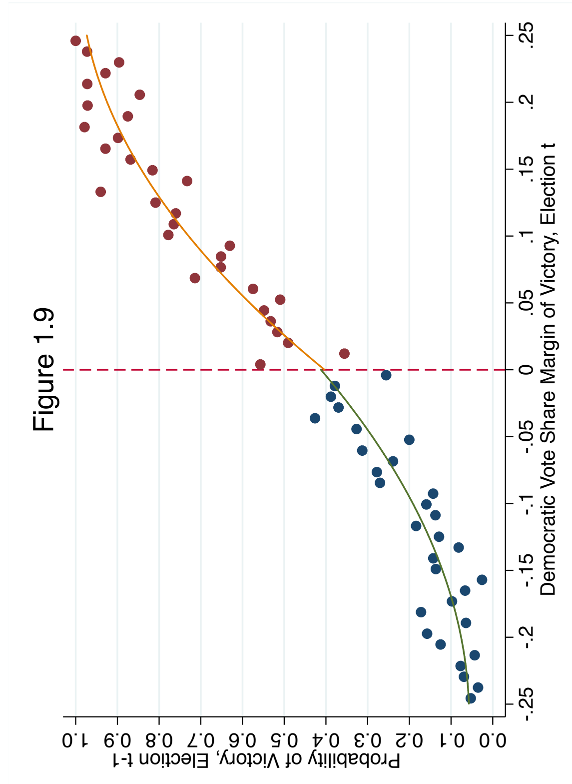


Figure 1.9

Table 1.4: Estimated Coefficient on Incumbency Advantage Indicator from Regressions with a Range of Baseline Covariates

VARIABLES	(1) Democratic Vote Share of Margin, election t+1	(2) Democratic Vote Share of Margin, election t+1	(3) Democratic Vote Share of Margin, election t+1	(4) Democratic Vote Share of Margin, election t+1	(5) Democratic Vote Share of Margin, election t+1	(6) Residual Vote Share	(7) 1st Dif. Vote share, t+1	(8) Democratic Vote Share of Margin, t-1
Victory, election t	0.0775***	0.0784***	0.0775***	0.0774***	0.0782***	0.0748***	0.0788***	-0.000886
	(0.0113)	(0.0108)	(0.0114)	(0.0114)	(0.0108)	(0.00371)	(0.0133)	(0.0108)
Dem. Vote share, t-1		0.293***			0.298***			
		(0.0187)			(0.0187)			
Dem. win, t-1		-0.0166**			-0.00655		-0.175***	0.240***
		(0.00697)			(0.00740)		(0.00967)	(0.00754)
Dem. political experience			-0.00112*		-0.000156		-0.00184	0.00240
			(0.000594)		(0.00335)		(0.00425)	(0.00270)
Opposition political experience			0.000605		-0.000226		-0.00782*	0.0108***
			(0.000930)		(0.00373)		(0.00429)	(0.00295)
Democrat Electoral experience				-0.00122**	-0.00292		-0.00306	0.000196
				(0.000578)	(0.00326)		(0.00414)	(0.00259)
Opposition electoral experience				0.000773	0.00332		0.0113***	-0.0113***
				(0.000892)	(0.00355)		(0.00410)	(0.00282)
Constant	0.453***	0.319***	0.454***	0.454***	0.313***	-0.0435***	0.0436***	0.383***
	(0.00797)	(0.0113)	(0.00809)	(0.00812)	(0.0116)	(0.00278)	(0.0106)	(0.00852)
Observations	6,559	6,559	6,559	6,559	6,559	6,559	6,559	6,559
R-squared	0.684	0.710	0.684	0.684	0.712	0.058	0.127	0.726

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: All regressions include polynomials of the runnning variable, Democratic Vote Share of Margin in election t, up to the fourth degree and their interaction with the Victory t variable.

1.13. Conduct a parallel RDD analysis on the baseline covariates from 1.12 (e.g. demshareprev and othofficeexp). What would you expect? What are the results?

We expected the baseline covariates such as demshareprev and othofficexp to be insignificant when we ran the parallel RDD analysis. As you can see in the above rdrobust output, all the baseline covariates are insignificant. This is to be expected as the treatment (winning) should not affect predetermined characteristics.

Table 1.5: Parallel RDD Analysis on Baseline Covariates

VARIABLES	(1) Dem. Vote share, t-1	(2) Dem. win, t-1	(3) Dem. political experience	(4) Opposition political experience	(5) Democrat Electoral experience	(6) Opposition electoral experience
RD Estimate	0.000884	0.0414	0.221	0.134	0.219	0.118
	(0.0121)	(0.0500)	(0.236)	(0.201)	(0.241)	(0.203)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Part 2: Regression Discontinuity Estimates of Class Size Effects

2. The file angrist_lavy.dta contains replication data for the 5th grade class size results reported in Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. Quarterly Journal of Economics, 114(2), 533-575. Henceforth A&L. You can find variable definitions using the describe command in STATA. Read the article by Angrist & Lavy (1999) prior to conducting the analysis.

2.1. What is the key causal question that A&L are attempting to answer? Why is the answer to this question interesting from a policy perspective?

The causal question which Angrist and Lavy (A&L) are attempting to answer is: What is the effect of class size on school achievement (proxied by test scores). Parents and teachers generally support smaller class sizes due to the belief that it improves academic performance. From a policy perspective, class size as an "input" is far easier to manipulate and measure than other factors that might improve outcomes (e.g. teacher quality). Less both financial and spatial constraints, manipulating class size is relatively simple. Due to this fact in combination with parental preference, class size is often at the forefront of many academic outcome discussions and policy goals.

There are several reasons why understanding the relationship between class size and school achievement is important when constructing education policy. Large gains in achievement with smaller class size would encourage policies that cap class size. Because the relationship of achievement and class size may be non-monotonic, for instance achievement levels at class sizes of 15 and 20 may not differ as greatly as class sizes of 20 and 25, 30 and 35, etc., it is important to determine the relationship between class size and achievement to optimize performance levels while balancing costs with these benefits.

Experiments exploring the relationship between class size and achievement are difficult to conduct because of both financial and ethical constraints. Randomized experiments are expensive, and most parents may be reluctant to keep children in the experiment if they were randomly selected into a large class size, although an experiment testing this has been done (Tennessee's Project Star). In addition, as explained below, a multitude of confounding factors influence school achievement, which may indicate other solutions such as magnet or charter schools to close the achievement gap and increase educational performance.

2.2. Explain why the conditional independence assumption is unlikely to hold in this case. Specifically, why is a simple regression of student achievement on class size and controls unlikely to yield a causal estimate of the impact of class size on achievement?

The Conditional Independence Assumption (CIA): is a condition when under a particular treatment selection bias disappears. In other words, when focusing on individuals with the "same" characteristics, the only difference between such individuals is the treatment; essentially creating a randomized experiment, where treatment status is as good as randomly assigned. Due to such randomization, uncontrolled variables will not bias results and a OLS regression, including controls, would suffice.

When focusing on the impact of class size on student achievement, there is no way for CIA to hold. Running a simple regression with controls would be insufficient because omitted variables, related to both achievement (dependent variable) and class size (independent variable) would bias the the impact of class size on achievement (up or down). Socioeconomic status is one example of a variable that is correlated with both larger class sizes and decreased levels of academic achievement.

A simple regression model will not yield a causal estimate because the variation in class size may covary with other factors in the model. Maimonides' Rule can be used as a potentially exogenous source of variation in class size, as an instrument to estimate the effect of class size (A & L, 1999, p. 535). Maimonides' Rule 'z' affects the outcome of interest 'y,' (test scores as a measure of achievement) only through its relationship to 'x,' (class size). The rule states that class size must be limited to 40, splitting a group of 41 into two, thus creating natural discontinuity points. Angrist & Lavy (1999) use this rule in a regression discontinuity design to determine a causal relationship of class size on achievement.

2.3. A&L attempt to derive causal estimates of the effect of class size on student achievement using a fuzzy regression discontinuity design. Briefly discuss what is meant by a fuzzy RD design. Explain what Maimonides' Rule is and why it allows for the estimation of causal effects. Explain why the current application is a fuzzy RD rather than a sharp RD.

Israel's public schools practice Maimonides' Rule, where two teachers are appointed once a class has more than 40 students. Once a cohort reaches increments of 40+1 students, an additional class is created. This produces different mean class sizes at each threshold, yet always results in smaller classes at each threshold, creating discontinuities in class size at each increment. Maimonides' Rule can be used to instrumentalize class size through its effect on class size. Because the rule is an exogenous source of variation, it can be used to causally estimate the effect of class size on academic achievement. This is because it does not covary with other variables that typically affect academic outcomes.

A Sharp Regression Discontinuity (RD) Design is when a treatment threshold is absolute. In this case once a threshold is reached, the treatment will happen. There are never situations where individuals who don't reach the threshold get treatment and there are never individuals who reach the threshold that don't get treatment. It is worth noting that a Sharp RD is closely associated with a Randomized Experiment. A Fuzzy RD is different because there is a change in the probability of treatment. Not a change in the rule itself, in this case Maimonides' Rule does not change. The treatment status is fuzzy, not the cutpoint. A Fuzzy RD is closely associated with Instrumental Variable Design. In Sharp RD, the jump in the outcome at the cutpoint is the estimate of causal impact of the treatment; in Fuzzy RD, the jump in outcome is divided by the jump in probability of treatment (in this case, enrollment indicates the probability as it is forcing the discontinuities implied by the rule and treatment is indicated by class sizes of increments of 40+1) (class notes, p.12). While there is a small potential for parents to manipulate their child's class size by self-selecting into a school district with lower enrollment (closer to 80 than 120 or 140 will produce smaller average class sizes), this is an unlikely scenario. In addition, evidence indicates that private school attendance is extremely rare in Israel, indicating that estimates based on Maimonides' Rule can be considered valid as they are not subject to manipulation.

2.4. Create a figure similar to Figure 1 (fifth grade) in A&L. Note you will need to use STATAs twoway scatter command to create this figure. You will also have to first use STATAs egen command to create "bins" of class size. Create these bins by taking the average of class size by total enrollment in grade (c_size). Name this new variable msize. Your graph should show both Maimonides' Rule and msize as a function of c_size. Place lines in your figure using the xline option at the discontinuities implied by Maimonides' Rule (i.e. at 41, 81, 121, and 161). Discuss what the figure you created reveals.

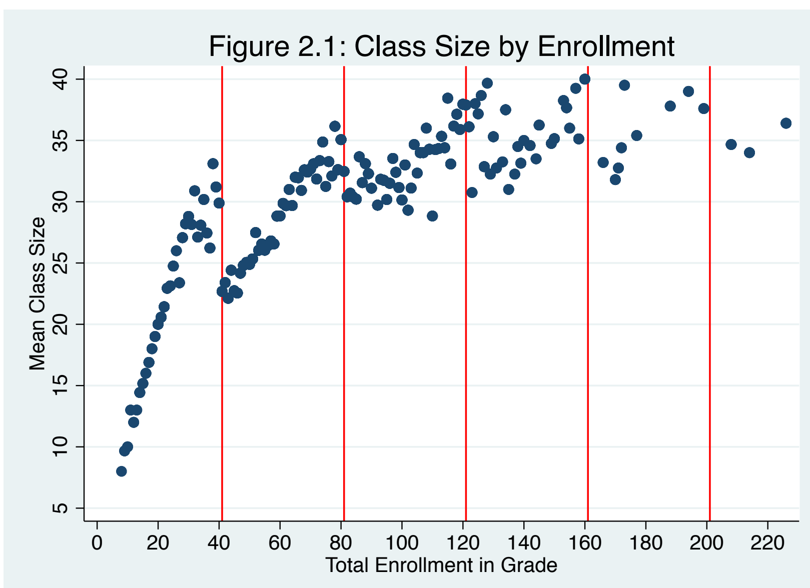


Figure 2.1 suggests that Israeli public schools, more specifically fifth grade classes, follow Maimonides' rule. The red vertical lines represent critical intervals of Maimonides' Rule, where fifth grade cohorts peak at 40 students per class and sharply dip after enrollments in increments of 40 are reached. Please note that these vertical lines are strategically placed at 41, 81, 121, and

161 where the rule creates a natural discontinuity cutpoint. We observe the first two intervals (enrollment of 41, 81) produce sharper results than the next three (121, 141, 201). This is because the overall magnitude of class size reduction lessens at each threshold. For instance, a cohort of 41 splits to produce a mean class size of 20.5, while a cohort of 81 produces a mean class size of 27, 121 produces 30.25, and so on. This is observed in the figure, as the sharpness of the results decrease with each cutpoint created by Maimonides' Rule. This lends further credence into the assumption that this is a Fuzzy RD because knowledgeable parents could locate to school districts with smaller cohorts. These smaller cohorts do not violate Maimonides' Rule, but create the potential for selection into smaller classes based on district total enrollment.

2.5. Now let's look more closely at the first discontinuity which occurs at an enrollment of 41. Specifically, use STATAs cmogram command to create a graph of the relationship between class size and enrollment. Limit enrollment to the range of 1 through 80. Use the cut() option to note that the discontinuity occurs at 40. Use the line() option to place a vertical line at the discontinuity at 40. Add a quadratic trend using the qfit option. Discuss the figure you created.

Figure 2.2: Class Size by Enrollment, Limited to 80

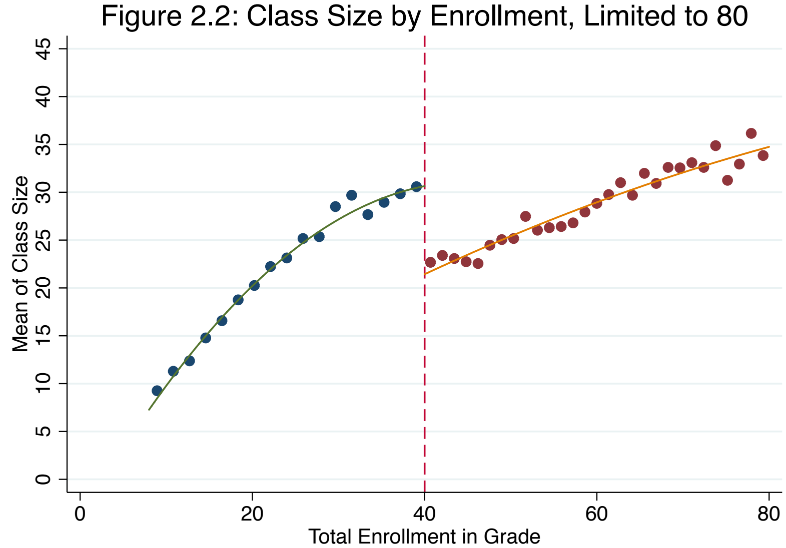
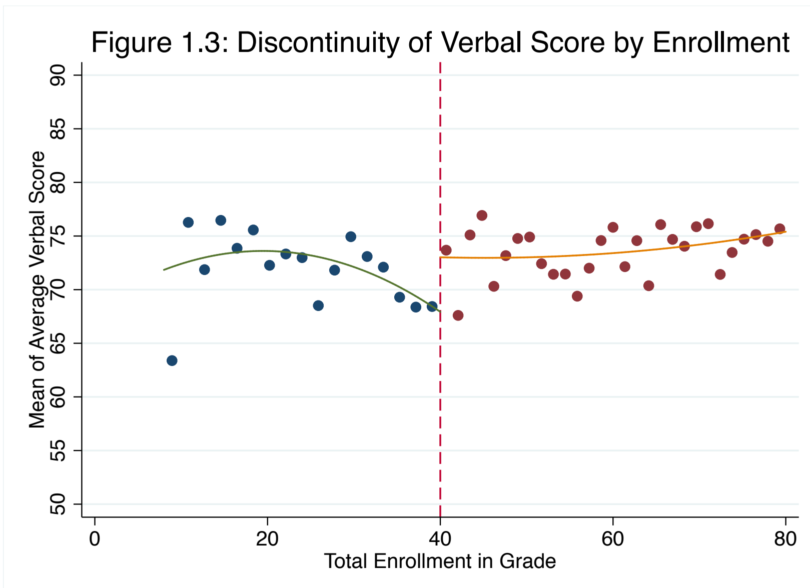
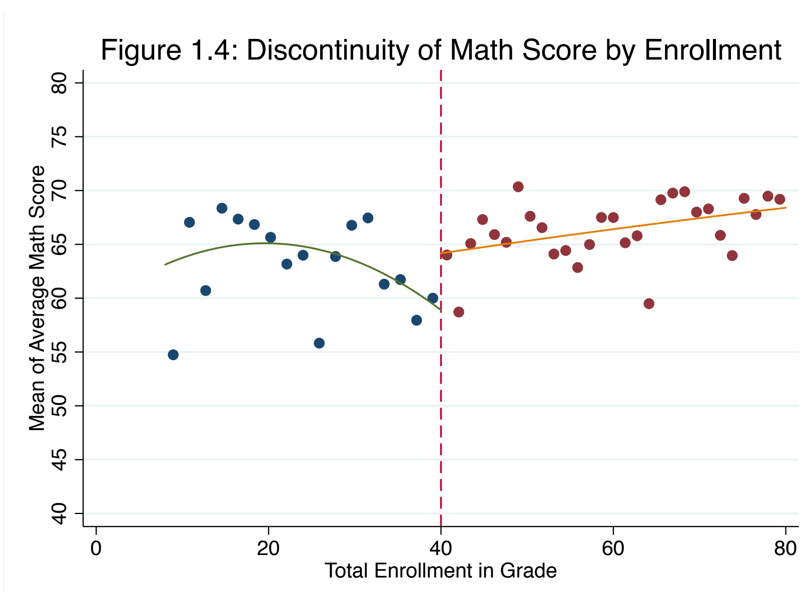


Figure 2.2 depicts the relationship between enrollment (x-axis) and average class size (y-axis) following Maimonides' Rule. It indicates that Maimonides' Rule does affect class size, and more specifically, indicates the discontinuity created by Maimonides' Rule when enrollment reaches 40 students. There is no doubt once the total enrollment exceeded 40, average class size decreased as a result of following Maimonides' Rule. Without the rule, we assume the monotonic trend (indicated by the color blue before the cutpoint) would continue. In American schools,

there would be different “cutpoints” in different municipalities, based on ability to pay. Israel offers a unique look into an exogenous source of variation in class size by applying a consistent rule or cutpoint.

2.6. See if you can detect a discontinuity in average verbal and average math scores around 40 by creating figures similar to the ones you created in part 2.5 but replacing class size with average verbal and average math scores. Discuss the figures you create. Do you see any evidence of a discontinuity?





In Figure 1.3, the dependent variable is average verbal scores and the independent variable is total enrollment in the grade. Again we observe a discontinuity when total enrollment is greater than 40, as implied by Maimonides' Rule and depicted by the vertical dashed line. As we can see, before the cutpoint (shown in blue), average verbal scores start to decrease when approaching an enrollment of 40 students and sharply increase (by approximately 6 points) after the enrollment reaches 40 (shown in red) and the class is split following Maimonides' Rule. The increase in mean verbal scores after the cutpoint produces similar results as a class of 20 before the cutpoint. This is because the the rule indicates that following an enrollment of 41, the class is split into two classes of 20.5. It should be noted, however, that when the cohort size continues to grow closer to 80, we would expect a similar dip in test scores because class sizes are becoming larger. We do not observe this similar trend approaching an enrollment of 80 students.

In Figure 1.4, the dependent variable is average math scores and the independent variable is total enrollment in the grade. The same discontinuity is observed at the cutpoint, with a sharp increase of approximately 5 points. The scores after the cutpoint are again similar to scores of enrollment at 20. As enrollment approaches 40, we observe a large dip in math scores, implying that large class size leads to lower scores; however, once again, we do not observe the same trend of decreased scores as an enrollment of 80 is approached.

2.7. Estimate the effect of class size on math scores using OLS without any controls, and then by sequentially adding the percentage of disadvantaged students in the class and enrollment as controls. These are Table 2 of A&L. Create a table of your results and discuss them. Are the results consistent with the notion that students perform better if they are in smaller classes?

Table 2.1: OLS Estimate of Class Size on Math Scores

	Average Math Score	Average Math Score	Average Math Score
Class Size	0.322***	0.0758**	0.0185
	(0.032)	(0.030)	(0.037)
Percent Disadvantaged Students		-0.340***	-0.332***
		(0.015)	(0.015)
Enrollment			0.017***
			(0.006)
Constant	57.66***	69.81***	70.09***
	(0.977)	(1.015)	(1.018)
Observations	2,018	2,018	2,018
R-squared	0.048	0.248	0.251

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In Table 2.1 above, which is an OLS model without controls, we observe a positive relationship between class size and math scores in column 1. More specifically, as class size increases by one student, we would expect a .322 unit increase in math scores, all else equal. Not only is this coefficient significant at the 1 percent level, the sign itself is the opposite of what conventional wisdom suggests. Typically we would expect that an increase in class size would decrease math scores on average. As shown in question 2.2, a simple regression is not adequate for an accurate representation of the relationship between class size and achievement. This is due to variations that are correlated with both the dependent and independent variables. This correlation causes bias, and violates the assumption of CIA. Finally it is important to note that there may be an inherent difference among Israeli students and findings may not be generalizable to other populations.

In columns 2 and 3 of Figure 2.1, once controlling for the variables of socioeconomic status and enrollment, the coefficient on class size not only shrinks in magnitude but loses its significance level. More importantly, the coefficient on socioeconomic status outweighs the impact of class size and it has a negative effect on math scores. This is not surprising because one would expect as the amount of poor students increased, average achievement would decline. Finally, although the coefficient is slightly smaller on enrollment, the coefficient on this variable is significant at all conventional levels, suggesting its impact on math scores is more viable than class size. This is likely due to the effect of Israeli schools implementing Maimonides' Rule.

2.8. Now limit your sample to enrollments of less than 81 (i.e. the sample on either side of the first discontinuity). Use this sample for parts 2.8 – 2.14. Re-center the forcing variable (c_size) so that it equals zero at the first discontinuity (i.e. subtract 41 from c_size). Create an indicator for enrollment greater than 40 since this is where the discontinuity occurs. Name this indicator small. Create a quadratic in the re-centered running variable and interactions between the running variable and the small indicator.

Please see attached do file.

2.9. Estimate the first stage results: the impact of Maimonides' Rule on class size. Run two regressions. In the first only control for the fraction of disadvantaged students and the level of the running variable. In the second, add the quadratic and interactions or the running variable. Create a table of your results and discuss them.

As seen in Table 2.2 below, the two models produce similar results for the estimate of the first stage. Model 1 produces a first stage of .5369 while Model 2 (which includes a quadratic and an interaction term for the running variable) produces a first stage of .5639. The F-Statistic for both models is substantially higher than 10, indicating the power of Maimonides' Rule as a valid instrument of class size. The F-Statistics for Models 1 and 2 are 462 and 284, respectively.

Table 2.2: First Stage Estimate of Maimonides' Rule

Model 1	Model 2
0.5369***	0.5639***
(0.0295)	(0.0301)
F(3, 1175) = 462.30	F(5, 1173) = 284.34

Note: The first model includes indicators for the fraction of disadvantaged students, and a recentered running variable. The second model includes the fraction of disadvantaged students, a recentered running variable, a quadratic term and an interaction of the running variable and small indicator.

2.10. Now estimate the reduced form results: the impact of Maimonides' Rule on outcomes, where the outcomes are average math and average verbal scores. Use the same two specifications as in part 2.9 and add the results to your table. Discuss the reduced form results.

Table 2.3: First Stage and Reduced Form Estimates of Maimonides' Rule

Model 1			Model 2		
FS Estimate	Reduced Form Math	Reduced Form Verbal	FS Estimate	Reduced Form Math	Reduced Form Verbal
0.5369***	-0.1709***	-0.1452***	0.5639***	-0.1762***	-0.1504***
(0.0295)	(0.0574)	(0.0434)	(0.0301)	(0.0588)	(0.0445)
F(3, 1175) = 462.30			F(5, 1173) = 284.34		

Note: The first model includes indicators for the fraction of disadvantaged students, and a recentered running variable. The second model includes the fraction of disadvantaged students, a recentered running variable, a quadratic term and an interaction of the running variable and small indicator.

The reduced form results can be seen in the graph above. Reduced form estimates of Maimonides' Rule explain the effect of the instrument on the outcome of interest. Coupled with the first stage that shows the rule effectively instrumentalizes class size, we can use the rule as an exogenous indicator of variation in class size to estimate academic outcomes. When adding additional covariates including squaring the running variable and adding an interaction on the running variable the estimates are similar – the reduced form increases only slightly.

2.11. Calculate the WALD estimates of the effect of class size on student achievement. Report these estimates in your table and interpret the results. Make sure you explain how the WALD estimates are derived.

Table 2.4, Panel A: First Stage, Reduced Form and WALD Estimates of Maimonides' Rule on Math and Verbal Scores

Model 1				
FS Estimate	Reduced Form Math	WALD Math	Reduced Form Verbal	WALD Verbal
0.5369***	-0.1709***	-0.31834	-0.1452***	-0.27051
(0.0295)	(0.0574)		(0.0434)	
F(3, 1175) = 462.30				

Note: This model includes indicators for the fraction of disadvantaged students, and a recentered running variable.

Table 2.4, Panel B: First Stage, Reduced Form and WALD Estimates of Maimonides' Rule on Math and Verbal Scores

Model 2				
FS Estimate	Reduced Form Math	WALD Math	Reduced Form Verbal	WALD Verbal
0.5639***	-0.1762***	-0.31252	-0.1504***	-0.26676
(0.0301)	(0.0588)		(0.0445)	
F(5, 1173) = 284.34				

Note: This model includes the fraction of disadvantaged students, a recentered running variable, a quadratic term and an interaction of the running variable and small indicator.

WALD estimates are calculated by dividing the reduced form by the first stage. The difference between the models is small, with Model 2 estimates being slightly smallerWhen adding controls between the models, the Math and Verbal Models 2 are slightly smaller. The interpretation of WALD estimates are as follows:

- Being in larger class size results in a .318 lower test score on Math in Model 1 when comparing the difference in means.
- Being in larger class size results in a .313 lower test score on Math in Model 2 when comparing the difference in means.
- Being in larger class size results in a .271 lower test score on Verbal in Model 1 when comparing the difference in means.
- Being in larger class size results in a .267 lower test score on Verbal in Model 2 when comparing the difference in means.

2.12. Now estimate the same two models you did before using 2SLS and create a table of your results. How do the 2SLS results compare to the WALD estimates?

When computing the 2SLS figures, which can be seen above, the numbers match our previous WALD estimates exactly, as they should. The only difference is the standard errors are included in 2SLS estimates. The coefficients on class size (using the rule as an instrument) are all significant at 99%.

Table 2.5: 2SLS Estimates Using Maimonides' Rule

	Average Math Scores		Average Verbal Scores	
	Model 1	Model 2	Model 1	Model 2
Class Size	-0.318***	-0.313***	-0.271***	-0.267***
	(0.121)	(0.118)	(0.098)	(0.098)
Percent Disadvantaged Students	-0.336***	-0.335***	-0.359***	-0.358***
	(0.018)	(0.018)	(0.015)	(0.015)
Running Variable	0.083***	-0.065	0.029	-0.007
	(0.031)	(0.089)	(0.025)	(0.066)
Running Variable Squared		-0.003		-0.001
		(0.002)		(0.002)
Interaction Variable		0.290*		0.073
		(0.176)		(0.133)
Constant	79.44***	78.12***	86.72***	86.36***
	(3.106)	(3.031)	(2.540)	(2.467)
Observations	1,179	1,179	1,179	1,179
R-squared	0.221	0.224	0.364	0.365

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Model 1 includes an indicator for percent of disadvantaged student and a running variable which is the enrollment variable recentered at the first discontinuity. Model 2 adds to these a squared running variable, and a variable that represents the interaction between the running variable and a discontinuity indicator.

2.13. Now let's vary the bandwidth using the model that only controls for the fraction of disadvantaged students and the level of the running variable. Create a table that shows the 2SLS results based on bandwidths of: +/- 20, +/-10 and +/-5 around the cutpoint. Create a table of your results and discuss. What happens to the estimates on the small indicator as you narrow the bandwidth? What happens to the standard errors and why?

Table 2.6: 2SLS Estimates Using the Small Indicator

	Average Math Scores			Average Verbal Scores		
	Using Bandwidth 20	Using Bandwidth 10	Using Bandwidth 5	Using Bandwidth 20	Using Bandwidth 10	Using Bandwidth 5
Class Size	-0.354**	-0.229	-0.406	-0.432***	-0.562***	-0.733***
	(0.160)	(0.235)	(0.266)	(0.127)	(0.165)	(0.183)
Percent Disadvantaged Students	-0.357***	-0.332***	-0.334***	-0.376***	-0.384***	-0.408***
	(0.023)	(0.039)	(0.056)	(0.017)	(0.027)	(0.041)
Running Variable	0.074**	0.287	0.321	-0.007	0.012	-0.214
	(0.0326)	(0.177)	(0.410)	(0.028)	(0.093)	(0.256)
Constant	80.65***	76.54***	79.79***	91.33***	94.77***	99.47***
	(4.280)	(6.753)	(7.867)	(3.336)	(4.569)	(5.174)
Observations	667	315	151	667	315	151
R-squared	0.253	0.260	0.221	0.361	0.265	0.280

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As bandwidths are narrowed, both our estimates on class size and our standard errors get larger. It must be noted that when bandwidth is reduced for average math scores column 2 (bandwidth 10) seems to be an anomaly because our estimates decreased slightly rather than increased as expected. This may be due to finding the optimal bandwidth.

The reason why changing the bandwidth affects standard errors is because by working in a smaller range, or looking only data that is close to the discontinuity, we decrease the sample size. This does help by reducing the bias of estimators because the variable of interest are most similar when tighter to the center on each side of the discontinuity. In theory, this brings the experiment closer to CIA, and allows us to not to worry about additional variables. Yet, this creates a tradeoff between bias and standard error. As data is constricted using a narrower bandwidth around a discontinuity, our estimates will be less biased; however, limiting the data increases standard errors or overall preciseness.

2.14. Now use STATA's rd command to estimate local linear regressions of the effect of class size on verbal and math scores. Finally, do the same using STATA’s rdrobust command. Create a table of results and discuss/compare them to your other results.

Table 2.7: Estimates of Local Linear Effect of Class Size on Achievement

	Math			Verbal		
	Bandwidth 20	Bandwidth 10	Bandwidth 5	Bandwidth 20	Bandwidth 10	Bandwidth 5
Class Size	-0.444**	-0.253	-0.028	-0.506***	-0.362**	-0.218
	(0.192)	(0.166)	(0.176)	(0.187)	(0.174)	(0.206)
Observations	1,179	1,179	1,179	1,179	1,179	1,179

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.8: Estimates of Local Linear Effect of Class Size on Achievement Using Optimal Bandwidth

	Average Math Score	Average Verbal Score
RD Estimate of Class Size	-0.354	-0.644*
	(0.329)	(0.370)
Observations	288	352

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

When using the rdrobust command we see a similar coefficients to the results of local linear regression, concerning the impact of class size on verbal and math scores. More specifically when using the rdrobust command for math scores, we can see our coefficient lies between the bandwidths of 20 and 5 used in the local linear regression. When the rdrobust command is applied to verbal scores, it appears to produce a coefficient that has applied as an optimal level of a bandwidth greater than 20 that is less statistically significant. The coefficient is somewhat larger than the one produced by the 20 bandwidth.

As stated in the previous question, as bandwidth narrows bias is reduced, however standard errors increase. The narrowing of the bandwidth results in a reduction of statistical significance, observed in the local linear regression model (Table 2.7).When a bandwidth of 5 is applied, neither verbal nor math scores are statistically significant. The rdrobust command produced a nonsignificant coefficient for average math scores and the lowest significance (p<.01) for average verbal score.

Finally, the relationship of the magnitudes of the effect of math and verbal scores when bandwidth narrows are the exact opposite of the results in Table 2.6 (p. 23). As bandwidth narrowed in the 2SLS Estimates, we observed an increase on the impact of class size on math and verbal scores. In Table 2.7, we see the opposite trend – as bandwidth narrows there is a decrease in the overall magnitude of the impact of class size on math and verbal scores.

When comparing our rdrobust results to the results of 2SLS in question 2.13 (Table 2.6), our coefficients of class size and their impact on math and verbal scores fall within the range of coefficients produced by the different bandwidths.

2.15. Now let's once again use the full sample. Create a series of 5 indicator variables called small1-small5. These indicators should take the value of 1 if enrollment is on the right side of each of the discontinuities implied by Maimonides' Rule. Use these indicators as instruments to estimate models similar to column 3 and 9 of Table IV in A&L (i..e. models that include percent disadvantaged, enrollment and enrollment squared. Also estimate the first stage and show those results. What is the F-statistic for the excluded instruments? What does the F-stat suggest?

Table 2.9: 2SLS Estimates of the Effect of Class Size on Achievement Using the Small Indicator

	Average Math Score	Average Verbal Score
Class Size	-0.286**	-0.223***
	(0.113)	(0.083)
Percent Disadvantaged Students	-0.351***	-0.366***
	(0.016)	(0.012)
Running Variable	0.068**	0.004
	(0.032)	(0.024)
Running Variable Squared	-0.0001	7.16e-05
	(0.0001)	(8.97e-05)
Constant	76.39***	85.39***
	(2.187)	(1.604)
Observations	2,018	2,018
R-squared	0.226	0.350

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.10: Estimate of First Stage of Small Indicator

VARIABLES	Class Size
small1	-8.015***
	(0.745)
small2	-5.437***
	(0.509)
small3	-0.383
	(0.535)
small4	3.796***
	(1.045)
small5	7.862***
	(1.645)
poor	-0.0477***
	(0.00817)
enrollment	0.555***
	(0.0282)
enroll2	-0.00201***
	(0.000144)
Constant	11.75***
	(0.647)
Observations	2,018
R-squared	0.531

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

F(8, 2009) = 320.26

When running a joint significance test for the 5 small indicator variables, we produced an F-Statistic of (5, 2009) 32.59. A power of 10 or above indicates a valid instrument, so the F-Statistic, at more than three times that amount, indicates the validity of the small indicator as an instrument for class size.

2.16. Now limit your sample to what A&L call the discontinuity sample by creating an indicator for that sample and then only using observations for which your indicator equals one. Specifically, create an indicator for observations close to the discontinuities using the command:
gen disc= (c_size>=36 & c_size<=45) | (c_size>=76 & c_size<=85) | (c_size>=116 & c_size<=125)

Please see attached do file.

Estimate the same models as in part 2.15 using only observations in the discontinuity sample. Create a table of the results for parts 2.15 and 2.16 and discuss these results. How do they compare to the corresponding results reported in Table IV of A&L? Finally, re-estimate the models from parts 2.15 and 2.16 this time using by Maimonides' Rule (func1) as the instrument instead of the indicator variables. Add these results to your table. Do they replicate the A&L results in Table IV?

Table 2.11: Estimates of the Effect of Class Size on Achievement

	Full Sample		Discontinuity Sample	
	Average Math Scores	Average Verbal Scores	Average Math Scores	Average Verbal Scores
Class Size	-0.286**	-0.223***	-0.511*	-0.645**
	(0.113)	(0.0827)	(0.280)	(0.259)
Percent Disadvantaged Students	-0.351***	-0.366***	-0.425***	-0.461***
	(0.0162)	(0.0119)	(0.0462)	(0.0432)
Running Variable	0.0683**	0.00364	0.288**	0.137
	(0.0320)	(0.0235)	(0.122)	(0.108)
Running Variable Squared	-0.000117	7.16e-05	-0.00125**	-0.000472
	(0.000122)	(8.97e-05)	(0.000612)	(0.000509)
Constant	76.39***	85.39***	74.90***	93.36***
	(2.187)	(1.604)	(5.766)	(4.735)
Observations	2,018	2,018	471	471
R-squared	0.226	0.350	0.193	0.204

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

When using the discontinuity sample, we observe effects that are much larger in magnitude than when using the full sample. This is because limiting the data around the cutpoint produces a more dramatic discontinuity effect. For example, the coefficient on class size with respect to verbal in the full sample was -0.233; it is -0.645 using only the observations around the cutpoint.

Table 2.12: 2SLS Estimates of the Effect of Class Size on Achievement
Using Different Samples and Instruments

	Full Sample		Discontinuity Sample Using Small Indicator		Discontinuity Sample Using Maimonides' Rule	
	Average Math Scores	Average Verbal Scores	Average Math Scores	Average Verbal Scores	Average Math Scores	Average Verbal Scores
Class Size	-0.286**	-0.223***	-0.511*	-0.645**	-0.490**	-0.599***
	(0.113)	(0.0827)	(0.280)	(0.259)	(0.205)	(0.175)
Percent Disadvantaged	-0.351***	-0.366***	-0.425***	-0.461***	-0.423***	-0.457***
	(0.0162)	(0.0119)	(0.0462)	(0.0432)	(0.0431)	(0.0389)
Running Variable	0.0683**	0.00364	0.288**	0.137	0.284***	0.126
	(0.0320)	(0.0235)	(0.122)	(0.108)	(0.107)	(0.0908)
Running Variable Squared	-0.000117	7.16e-05	-0.00125**	-0.000472	-0.00124**	-0.000444
	(0.000122)	(8.97e-05)	(0.000612)	(0.000509)	(0.000581)	(0.000468)
Constant	76.39***	85.39***	74.90***	93.36***	74.51***	92.50***
	(2.187)	(1.604)	(5.766)	(4.735)	(5.012)	(3.690)
Observations	2,018	2,018	471	471	471	471
R-squared	0.226	0.350	0.193	0.204	0.201	0.232

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Our results are similar but not exact to A&L's Table IV. All of our estimates are larger in magnitude, except for Verbal scores using the full sample. They do, however, follow the trend that estimates using the discontinuity sample will be larger in magnitude.

As the results in Table 2.12 show, using Maimonides' Rule as an instrument produces estimates that are more statistically significant than the small indicator (overall). In addition, the estimates using Maimonides' Rule, while not an exact replica of A&L's Table IV, are closer to the original estimates found by the researchers. Our estimates using Maimonides' Rule and the discontinuity sample are slightly larger than those reported in Angriest & Lavy (2009). Again, the estimates follow the overall trend that estimates using the discontinuity sample will be larger in magnitude than estimates using the full sample.

Table 2.13: Estimates of First Stage of Instruments

Small Indicator		Maimonides' Rule	
	Class Size		Class Size
small1	-7.689***	mrule	0.342***
	(1.636)		(0.0496)
small2	-1.768*		
	(0.911)		
small3	1.238		
	(0.816)		
o.small4	-		
o.small5	-		
poor	-0.0707***	poor	-0.0638***
	(0.0243)		(0.0235)
enrollment	0.470***	enrollment	0.135**
	(0.0726)		(0.0630)
enroll2	-0.00188***	enroll2	-0.000301
	(0.000422)		(0.000356)
Constant	16.25***	Constant	12.40***
	(2.704)		(2.279)
Observations	471	Observations	471
R-squared	0.412	R-squared	0.438
F (6, 464) 70.27		F (4,466) 126.28	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results in Table 2.13 indicate that Maimonides' Rule is a stronger instrument than the small indicator. They are both valid instruments using the F-Statistic as a measurement, but the F-Statistic on the first stage of Maimonides' Rule is nearly double that of the one for the small indicator, showing it has more power to exogenously instrument class size.