Regression Discontinuity Design

Draws on Lee and Lemieux JEL 2010

Some motivating pictures:

RD methods can be described as “models” to infer the causal effect of a treatment when the probability of participation varies **discontinuously** as a function of one or more observable characteristics which are also related to the outcome of interest

Examples:

--Schools with score gains below a certain cutoff are subject to sanction under NCLB

--students with test scores above a certain level get a GED

--students arriving at school after Oct 15th do not count in school’s test scores

--houses on different side of street are in different school districts

--Impact of unionization on wages using union election results

--Impact of insurance on health access using Medicaid 65 yr old eligibility

--Maimonides Rule for class size

Figure 1: Health Insurance Coverage Rates by Age, 1992-2001 NHIS

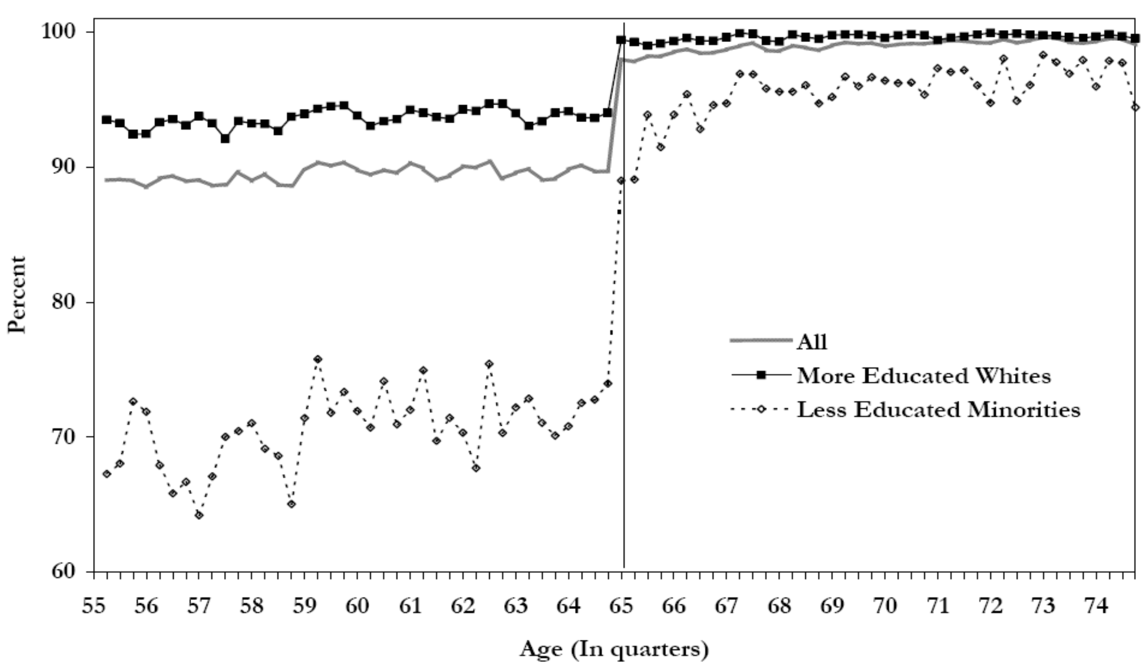
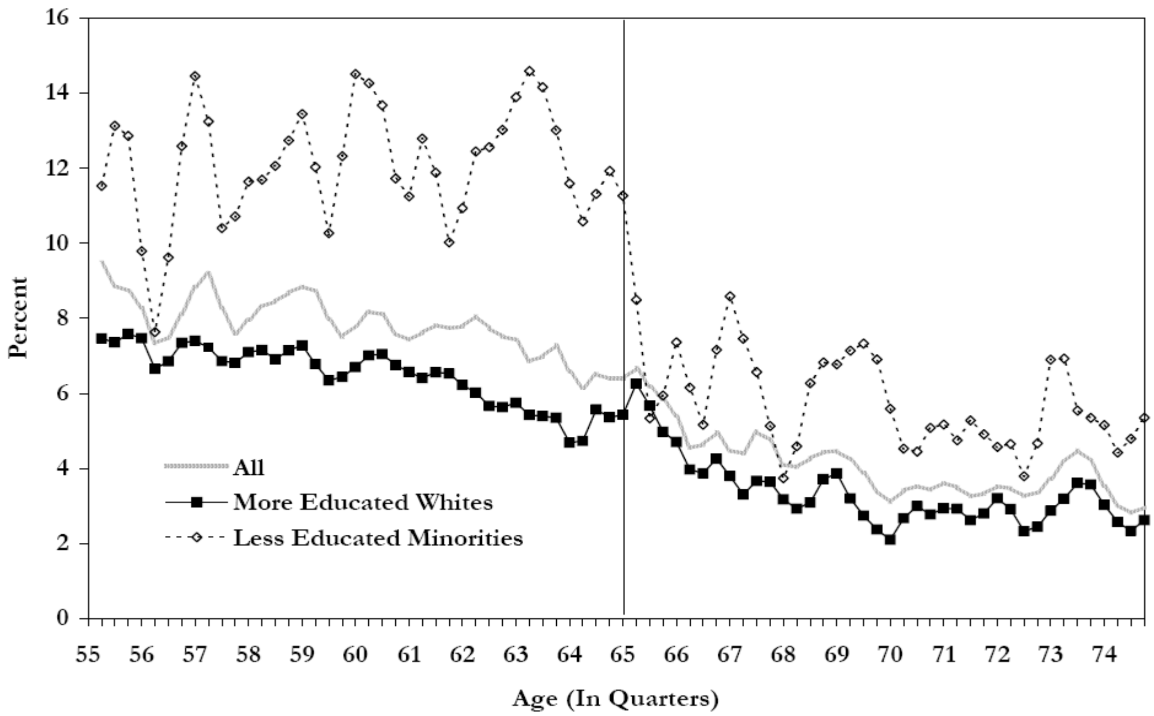


Figure 4: Percent Who Delayed Medical Care Last Year for Cost Reasons



Basic specification

$Y_i = \alpha + \tau D_i + \beta_1 X_i + e_i$

Where Di = 1 if Xi≥c

Where Di is the treatment dummy (0/1) and Xi is the assignment variable (e.g., student gain score)

Main Idea:

- Agents just below the cutoff are a good comparison group for agents just above the cutoff
- If individuals have *imprecise control* over assignment variable (test score), then treatment is random
    - If have ability to precisely manipulate assignment, then is invalid
- Much weaker than usual identification conditions. Do NOT need to assume have NO control over assignment variable (studying affects test scores, teacher behavior affect gain scores), but there is still an element of randomness
- As a result, the RD design can be analyzed like a random experiment

Assumptions of RD Model

Think about basic requirements for matching/selection on observables models—Identical treatment and control groups. What are implications?

1. **Random Assignment—imprecise control over assignment**

- Random assignment conditional on observables—this is like the exogeneity condition—all relevant factors are controlled for, there are no omitted variables correlated with treatment variable.

- Here this is satisfied in a trivial way—When $X_i \geq c$, D = 1. When $x<c$, D=0. As a result, if condition on X, there is no variation left in D, so it can't be correlated with any omitted variables.

- What is counterpart in this case? **Agents have imprecise control over assignment variable**. If have precise control, that is, if agents can manipulate X precisely, then the "experimental design" falls apart.

  Sometimes this is then the outcome economists look at
  "gaming" the system—e.g.,

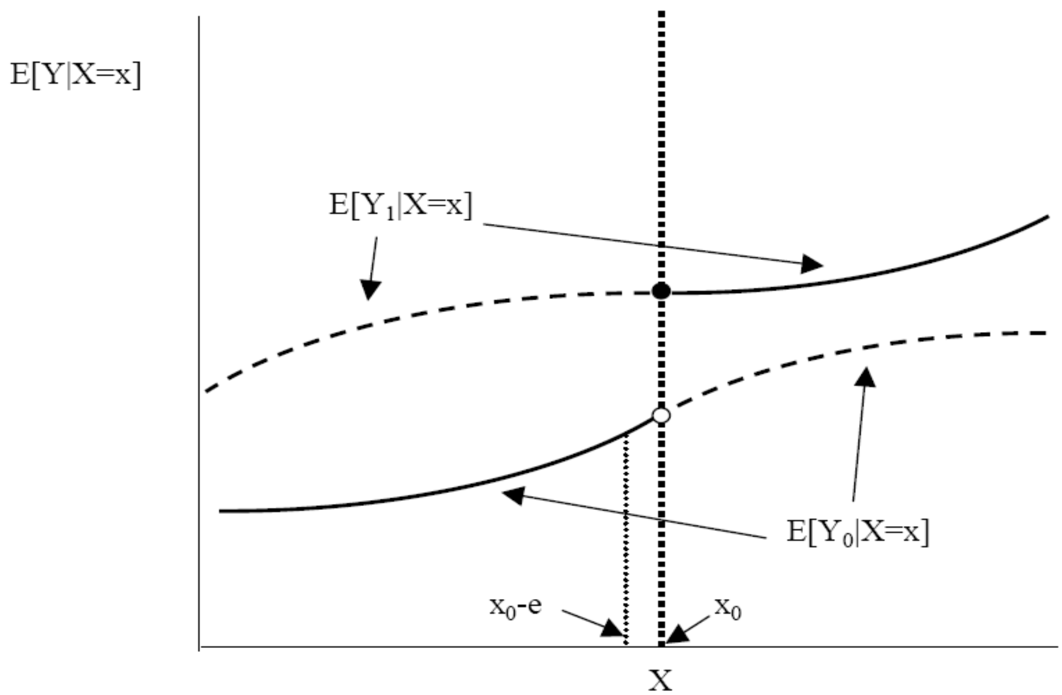
  Pg 292 "When there is a continuously distributed stochastic error component to the assignment variable—which can occur when optimizing agents do not have precise control over the assignment variable—then the variation in the treatment will be as good as randomized in a neighborhood around the discontinuity threshold."

2. **"Overlap"**—there are some individuals with same X and are assigned to either Treatment or Control
   Continuity

- Here this can never be satisfied, because there is a break in X that determined treatment

- Counterpart in this case is that All other factors are "continuous". Let Yi(1) be the outcome if an individual is treated and Yi(0) be the outcome for the same individual if they were not treated. Continuity means that E[Yi(1)|X] and E[Yi(0)|X] are continuous. Show figure 2 (alternative one from Evans here)

Figure 1: Regression Discontinuity, Continuous Covariate



- Lee and Lemieux argue that this continuity assumption is satisfied because of the random nature of the assignment

- Show Figure 3 to illustrate. Imagine treatment is assigned in this way. Y is employment. X is a random number assigned to you. Then everyone with assigned id above a certain number gets job training and below a certain number does not. Assigned number is irrelevant in regression. Lines are flat.

- Now imagine that a person was compensated for "bad draws" with money. High numbers get more money, low draft numbers get less money. We could think about a smooth relationship between money and employment, and then maybe the lines would have a positive slope. But people close to the cutoff still have roughly the same amount of money. Continuity is satisfied because there is not a discrete jump in effect of money at the cutoff point.

Testing for Imprecise Control and Continuity

1. **Learn a lot about assignment and how hard/easy it is to manipulate**

2. **Check distribution of X—is there a jump in distribution at the cutoff c? Check distribution of other covariates-- is there a jump in distribution at the cutoff c?**

   Example:
   Suppose X is test scores, and students with scores over 50% will win scholarships.

   Suppose have two types of students: Type A (more able and more aware of the assignment rule) and Type B (less able and less aware of assignment rule)

   Suppose students can precisely manipulate score—that is, suppose 50% of the questions are trivially easy and all students can get a score of 50% if check answers. Some randomness in type B students—some will pass and some fail. All type A should pass (only failures are type B).

   W = ability, awareness of rule
   X = score
   D = scholarship
   Y = earnings

   $Y_i = \alpha+\tau D_i+\delta_1 W_i+u_i$
   Where Di = 1 if $Xi \geq .50$
   $X = \delta_2 W_i + v_i$

So W is endogenous, except we observe X. No exclusion restrictions (δ_1, δ_2 can be zero or not), no assumptions about correlation between W, u, and v.

What does distribution of X look like, conditional on W =Type A, U=u?
With complete control over X, there will be truncated distribution for the type A above the cutoff. Draw as in figure 4.

What if test is such that Type A students cannot precisely generate a score of at least 50 percent? Can influence score through effort (imprecise control), but some Type A's will miss some questions due to randomness. Finish Figure 4.

If distribution looks like the one with imprecise control, then those who marginally failed and those who passed are otherwise comparable—can use RD design.

→Implication is that with imprecise control, the distribution of X is continuous. Without, will have a lumpy distribution of X in neighborhood of cutoff.

> Example: In Chile, have a privatized education system. Class size mandated at max of 30 students. Funding is on a per pupil basis. Class size paper looked at the discontinuity to examine effect of class size on student achievement. Later paper looked at class size itself. Big "lump" at 30 students. Suggests schools are reporting fewer students than have to keep from having to hire an additional teacher. But this effect is compounded by the school choice system. Since parents can observe class size, even if government officials cannot, implies that parents of kids in school with 2 classes of 15 may be different than parents in school with 1 class of 30. Results from previous paper not valid.

Restated: $Pr[W=w, U=u|X=x] = f(x|W=w, U=u)*Pr[W=w, U=u]/f(x)$ (Bayes rule)

→All observed and unobserved predetermined characteristics will have same distribution on either side of x=c in the limit, for smaller and smaller neighborhoods around c.

Show graphs for Dinardo and Lee below

3. Include other covariates and test for significance

Including covariates should not affect RD estimate—may lead to more precision, but if treatment is not precisely determined, should not change estimate.

4. When appropriate, perform RD regression on change in Y.

   If treatment is randomly assigned, lagged Y should not be related to D, and this should not affect RD estimate.

Dinardo and Lee Economic Impacts of New Unionization QJE paper

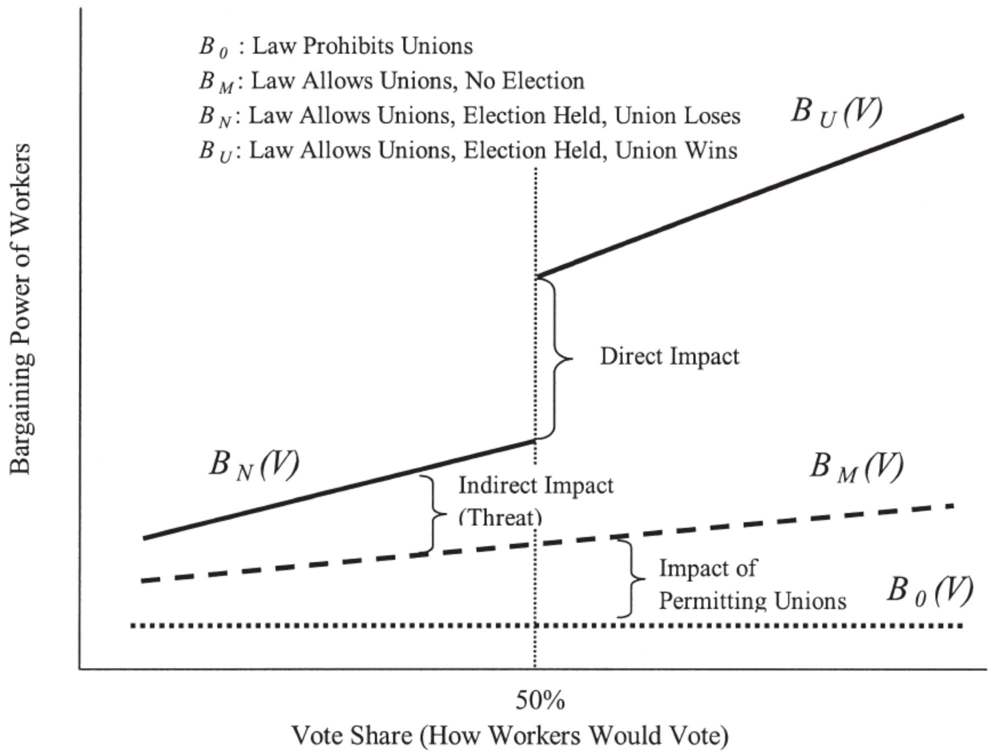


FIGURE I
Theoretical Relation between Employer Outcome and Vote Share

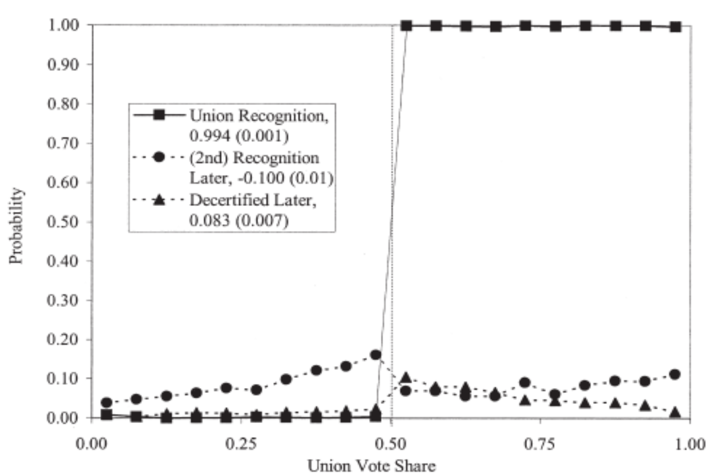


FIGURE IIIa
Recognition, Subsequent Certification or Decertification, by Union Vote Share.

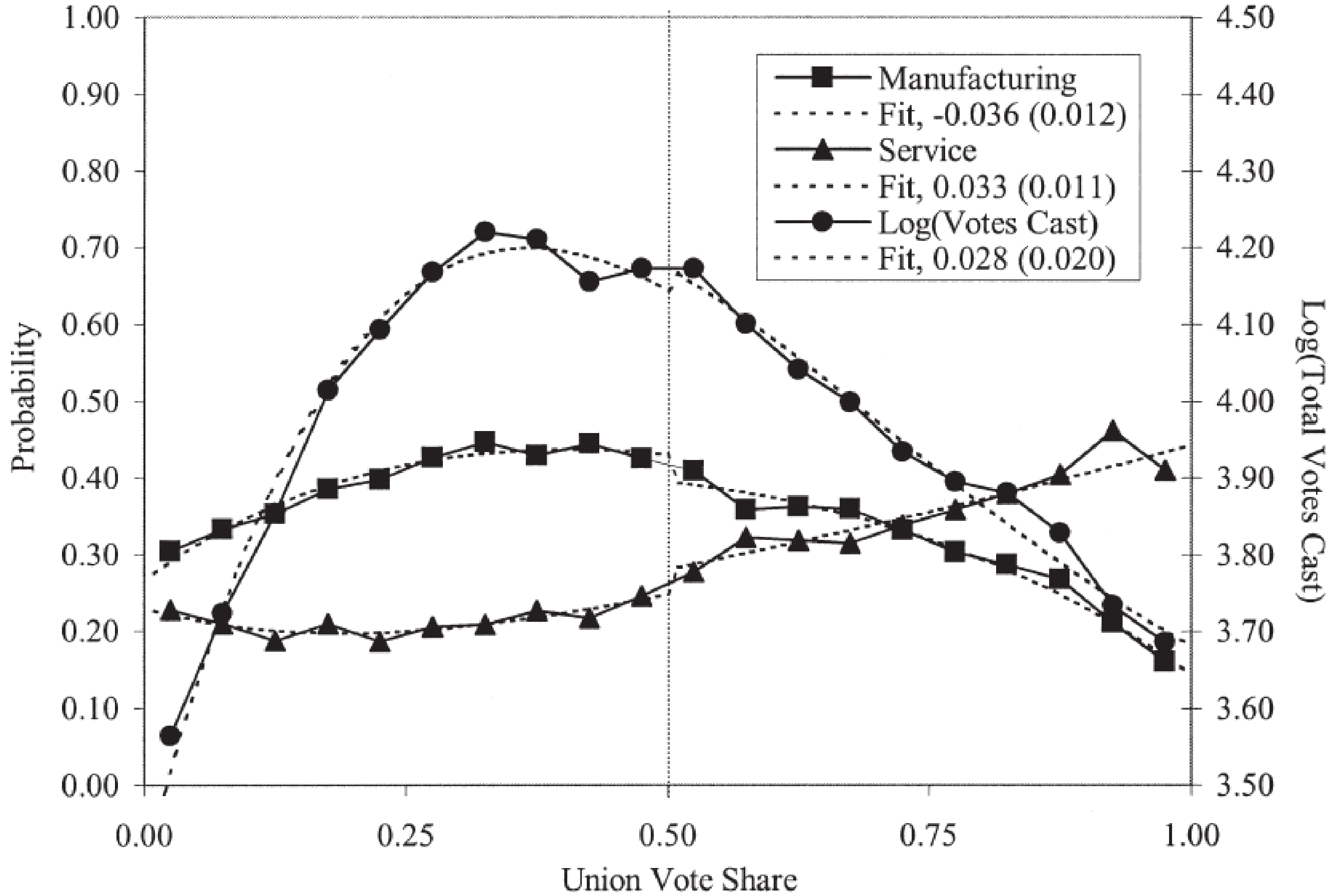


FIGURE VIa
Baseline Characteristics at Time of Election, by Union Vote Share

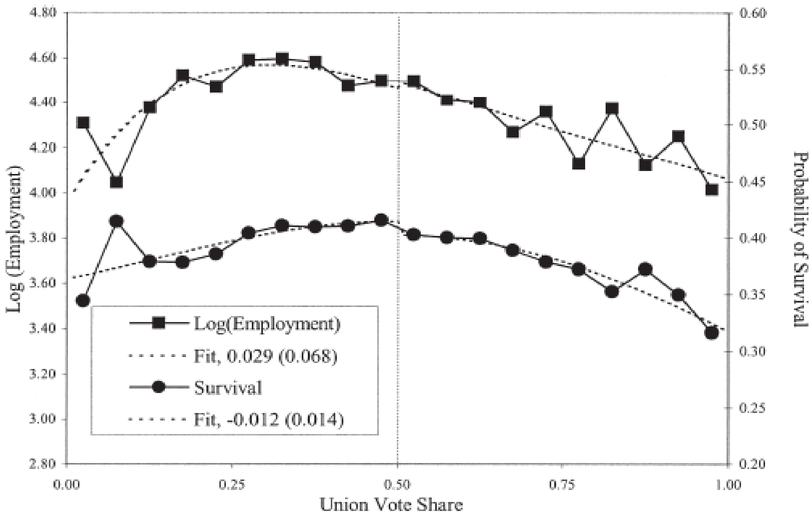


FIGURE IV

14

TABLE II
LEAST-SQUARES REGRESSION-DISCONTINUITY ESTIMATES OF UNION EFFECTS, LRD SAMPLE

Dependent variable	Coefficient on won election (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log(Hours)	0.009	−0.318	−0.260	−0.203	0.085	0.097	−0.024	0.015	0.018	0.028
	(0.087)	(0.036)	(0.063)	(0.063)	(0.080)	(0.080)	(0.056)	(0.051)	(0.050)	(0.049)
	[4733]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]
Log(Output)	0.079	−0.347	−0.293	−0.254	0.067	0.080	−0.043	−0.010	−0.004	0.011
	(0.094)	(0.042)	(0.072)	(0.073)	(0.090)	(0.091)	(0.055)	(0.050)	(0.050)	(0.049)
	[4730]	[28785]	[28785]	[28785]	[28785]	[28785]	[28785]	[28785]	[28785]	[28785]
Polynomial terms	0	0	1	2	3	4	4	4	4	4
Dependent variable	Level	Level	Level	Level	Level	Level	De-meaned	De-meaned	De-meaned	De-meaned
Include base mean?	No	No	No	No	No	No	No	Yes	Yes	Yes
Year dummies	No	No	No	No	No	No	No	No	Yes	Yes
Industry dummies	No	No	No	No	No	No	No	No	No	Yes

17

TABLE II
LEAST-SQUARES REGRESSION-DISCONTINUITY ESTIMATES OF UNION EFFECTS, LRD SAMPLE

Dependent variable	Coefficient on won election (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log(Wage)	0.015	−0.039	−0.041	−0.044	−0.005	−0.002	−0.026	−0.018	−0.018	−0.016
	(0.025)	(0.011)	(0.019)	(0.020)	(0.024)	(0.024)	(0.017)	(0.016)	(0.016)	(0.015)
	[4733]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]	[28796]
Polynomial terms	0	0	1	2	3	4	4	4	4	4
Dependent variable	Level	Level	Level	Level	Level	Level	De-meaned	De-meaned	De-meaned	De-meaned
Include base mean?	No	No	No	No	No	No	No	Yes	Yes	Yes
Year dummies	No	No	No	No	No	No	No	No	Yes	Yes
Industry dummies	No	No	No	No	No	No	No	No	No	Yes

Failure of Overlap condition gives rise to large issue of functional form

-- show graphs with Figure 1 with slope to it and the misspecified form with assumption of no slope. Also show nonlinear example

Note that this is a problem, even if you allow the slope to change.

- Because the effect is being measured using a break in the relationship, nonlinearity becomes a huge issue
- Another way to say this: consequences of incorrect functional form are much more serious here than in typical regression. In a linear regression where true functional form is nonlinear, estimated model can still be interpreted as a linear predictor that minimized specification errors—an average effect. Show a u shaped curve—linear model implies that average effect of x on y is zero—which is true.
- Because are minimizing specification errors globally, can still have large specification errors locally at cutoff→ large bias in RD estiamtes

Using a Polynomial

- A low order polynomial allows you to use all data, even data “far” from cutoff, which is more efficient.
- Common approach:
  1. Examine visual relationship—is there a visual break? And how many “flexion” points are there? (e.g., single flexion point indicates a quadratic)
  2. One approach: Go up an order of polynomial. E.g., if data appear to be linear, include at least a quadratic. Note that in doing this, need to include squares of X and cross products with X and D—X, D, X2, XD, X2*D
  3. By including higher order terms than are likely to be necessary, introduce inefficiency. Refine based on significance of coefficients, goodness of fit measures, and pattern of residuals.
  4. Alternative—test against unrestricted model

What order Polynomial? Testing Against unrestricted model
Example: suppose X is discrete

$Y_{ij} = D_j\beta_o + h(x_j) + \varepsilon_{ij}$

- x takes on discrete values $[x_1, x_2 \ldots x_j]$
- $x_k=0$ at discontinuity
- $D_j=1[x_j>0]$
- $h(x_j)$ polynomial in x
- This is a restricted model – parametric form of the regression with k parameters (k-1 in h and one for β_o)

Unrestricted model, run regression with complete set of dummies for x_j

- $Y_{ij} = \eta_j + \varepsilon_{ij}$
- J dummies in total

Goodness of fit test

$$G = \frac{(ESS_r - ESS_u)/(J-K)}{ESS_u/(N-J)}$$

- Under null that h(x) captures the time series characteristics, G is distributed as an F(J-K,N-K)
- J-K is difference in dof—number of restrictions
- N-J is dof of unrestricted model

See example with calculated F stat below—from Card Medicare example

Restricted model

N-1

```
. reg insured male white black hispanic _Ie* _Iyear* index index2 index_age65
> index2_age65 age65;

      Source |       SS       df       MS              Number of obs =   46950
-------------+------------------------------           F( 19, 46930) =  260.17
       Model |  334.010957    19  17.5795241           Prob > F      =  0.0000
    Residual |   3170.9949 46930  .067568611           R-squared     =  0.0953
-------------+------------------------------           Adj R-squared =  0.0949
       Total |  3505.00586 46949  .074655602           Root MSE      =  .25994
```

SSE_R

N-K-M

22

Unrestricted Model

```
. * run unrestricted model that has dummies for all quarters instead of polyno
> mial;
. * use for test in card/lee, equation (3);
. reg insured male white black hispanic _Ie* _Iyear* _Iage*;

      Source |       SS       df       MS              Number of obs =   46950
-------------+------------------------------           F( 93, 46856) =   54.01
       Model |   339.36356    93  3.64907053           Prob > F      =  0.0000
    Residual |   3165.6423 46856  .067561087           R-squared     =  0.0968
-------------+------------------------------           Adj R-squared =  0.0950
       Total |  3505.00586 46949  .074655602           Root MSE      =  .25993

------------------------------------------------------------------------------
     insured |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .0079442   .0024272     3.27   0.001     .0031869    .0127014
```

SSE_u

N-J-M

J+M-1
(no constant)

23

- SSEu = 3165.64
- N-J-M = 46856
- J+M = 93
- SSEr =3170.99
- K+M = 19
- J – K = 93-19 = 74
- F = 1.07
- = [(3170.99-3165.64)/74)/[3165.64/(46856)]
- P-value w/ 74 and 46,856 degrees of freedom is 0.31
- Cannot reject null that h(x) and full set of effects for x explain the same time pattern in y

So one approach to functional form problem is to "parameterize" the function by specifying a polynomial in X

Second method is to use non-parametric methods

- Nonparametric methods allow for flexible estimates of regression function—smooth out the data into a function without imposing a particular functional form (quadratic) on the data
- While polynomial functions are good, still are using all of the values of X to estimate the function—RD design depends on local estimates at the cutoff
- Basic idea is to specify a narrow bandwidth and only use data "close" to the cutoff" to estimate RD effect Indicate in graph
- "Kernal" estimators are basically weighted averages—can be a "rectangular kernel" which computes the average over a range (the bin). Also can have triangular kernals or other function forms of kernals where provide more weight to observations "close" to the point of interest
- However, draw Figure 2—note the problem with using a local average like this when there is a relationship between X and Y. B' – A' will overstate the true effect B-A

- Bigger bins will give more data to estimate average values, but lead to larger bias
- One way to handle this is to run local linear regressions inside the bins
- Another way is to run RD estimates with different bin sizes to show robustness of estimates
- Bottom line: lots of specification with different orders of polynomials or different bin sizes to show robust results.

Interpretation of RD Estimates (3.3)

→What if treatment has heterogenous effects? To what extent are the results generalizable?

One issue from IV and panel estimation was how generalizable the results are—what treatment effect does the estimator capture and can this be extended to the whole population?

e.g., if do RD approach with age 65 shift to Medicaid, does this identify the population effect of insurance on trips to the doctor? Is this the same effect would see for people with and without insurance (randomly assigned) at age 40?

Two potential answers:
--RD estimates are the treatment effect at the discontinuity threshold
--RD estimates represent a weighted average treatment effect across all individual

Start with simple case of estimate of effect of treatment when treatment has a uniform effect:

$Y_i = \alpha+\tau D_i+\beta_1 X_i+e_i$

Where $D_i = 1$ if $X_i \geq c$ (Draw Figure 2)

B-A = lim E[Yi|Xi=c+e] – limE[Yi|Xi = c+e]

=E[Yi(1) – Yi(0)|x=c]

Can do this inference because underlying functions E[Yi(1)|X] and E[Yi(0)|X] are continuous

Now repeat this with the test example where individuals have imprecise control:

Back to test example

W = ability, awareness of rule
X = score
D = scholarship
Y = earnings

$Y_i = \alpha+\tau D_i+\delta_1 W_i+u_i$
Where $D_i = 1$ if $X_i \geq c$
$X = \delta_2 W_i + v_i$

$B-A = \lim E[Y_i|X_i=c+e] - \lim E[Y_i|X_i = c+e]$

Now let τ vary for an individual with $W=w$ and $U = u$
$Y = D\tau(W,U) + W\delta 1 + U$

The discontinuity gap is a particular average across individuals

We have the term $f(c|W=w, U=u)/f(c)$, which keeps it from being a simple average.
What are these weights?
Weights are proportional to the ex ante likelihood that an individual's realization of X will be close to c

DK the similarity of these weights across people—only have one realization of X for each person and don't observe f
If weights are similar→RD estimate is similar to overall average treatment effect
If weights are varied→RS gap is different from average treatment effect

Comparison of RD, Matching Models, and IV

Show Figure 5 and walk through that 3.5

Final checklist

1. To assess the possibility of manipulation of the assignment variable, show it's distribution

   What are we checking for? "Heaping" at cutoff

2. Present main RD graph using binned local averages

   Balance "undersmoothing" where hard to see functional form and "oversmoothing" where implies less variance than is there

3. Graph a benchmark polynomial specification

4. Explore sensitivity of results to range of bandwidths and range of orders of polynomial

5. Conduct a parallel RD analysis with baseline covariates
   These shouldn't show breaks is assignment is valid

6. Explore sensitive of results to including baseline covariates
   a. These shouldn't affect estaimted RD gap
   b. If do, indicate potential sorting of assignment due to discontinuity in covariates at cutoff
   c. Increases in standard errors indicate misspecified functional form