Regression Discontinuity

Basic Setup

Lee's Interpretation

Examples

Nuts and Bolts

Graphical Methods

Analysis

Cross-Validation

Regression Discontinuity

October 3, 2012

Living on the Edge

Additive model of treatment

$$Y_i = Y_{0i} + \tau_i T_i, \text{ where } \tau_i = Y_{1i} - Y_{0i}$$

Living on the Edge

Additive model of treatment

$$Y_i = Y_{0i} + \tau_i T_i, \text{ where } \tau_i = Y_{1i} - Y_{0i}$$

Treatment is determined by X_i, and threshold x_0

$$T_i = I[X_i \geq x_0] \sim \text{sharply determined}$$

$$T_i = I[X_i + \epsilon_i \geq x_0] \sim \text{imperfectly determined}$$

Living on the Edge

Additive model of treatment

$$Y_i = Y_{0i} + \tau_i T_i, \text{ where } \tau_i = Y_{1i} - Y_{0i}$$

Treatment is determined by X_i, and threshold x_0

$$T_i = I[X_i \geq x_0] \sim \text{sharply determined}$$

$$T_i = I[X_i + \epsilon_i \geq x_0] \sim \text{imperfectly determined}$$

Local average treatment effect

$$\begin{aligned} \tau_{LATE} &= E[Y_{1i} - Y_{0i} | X_i = x_0] \\ &= E[Y_{1i}(x_0) - Y_{0i}(x_0)] \end{aligned}$$

Living on the Edge

Assumptions

Living on the Edge

Assumptions

1. $Y_{0i}, Y_{1i} \perp\!\!\!\perp T_i | X_i$
    - This assumption is trivially met, because conditional on X, there is no variation left in T, so it cannot be correlated with unobservables.

Basic Setup
Lee's Interpretation
Examples
Nuts and Bolts
Graphical Methods
Analysis
Cross-Validation

Living on the Edge

Assumptions

1. $Y_{0i}, Y_{1i} \perp\!\!\!\perp T_i | X_i$
   - This assumption is trivially met, because conditional on X, there is no variation left in T, so it cannot be correlated with unobservables.
2. $0 < \mathrm{P}(T_i = 1 | X_i = x) < 1$
   - In the sharp regression design, this is *always* violated.

Living on the Edge

Assumptions

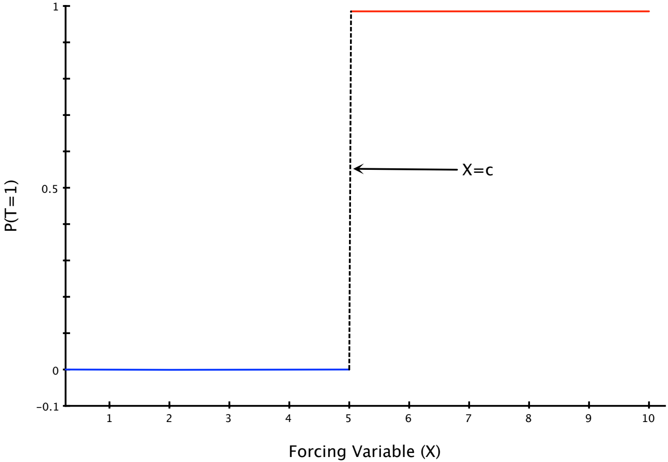
1. $Y_{0i}, Y_{1i} \perp\!\!\!\perp T_i | X_i$
   - This assumption is trivially met, because conditional on X, there is no variation left in T, so it cannot be correlated with unobservables.
2. $0 < \mathrm{P}(T_i = 1 | X_i = x) < 1$
   - In the sharp regression design, this is *always* violated.
3. $E[Y_0 | X = x]$ and $E[Y_1 | X = x]$ are continuous in x.
   - Continuity is required to compensate for the failure of the common support condition.

Living on the Edge

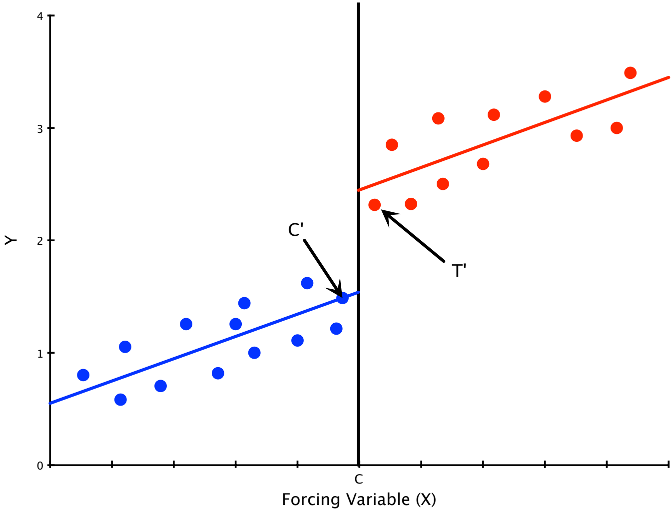
Assumptions

1. $Y_{0i}, Y_{1i} \perp\!\!\!\perp T_i | X_i$
   - This assumption is trivially met, because conditional on X, there is no variation left in T, so it cannot be correlated with unobservables.
2. $0 < \mathrm{P}(T_i = 1 | X_i = x) < 1$
   - In the sharp regression design, this is *always* violated.
3. $E[Y_0 | X = x]$ and $E[Y_1 | X = x]$ are continuous in x.
   - Continuity is required to compensate for the failure of the common support condition.
4. SUTVA

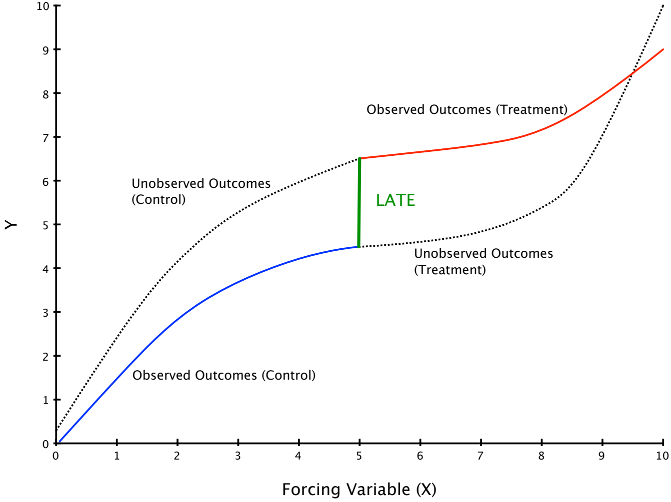
Probability of Treatment



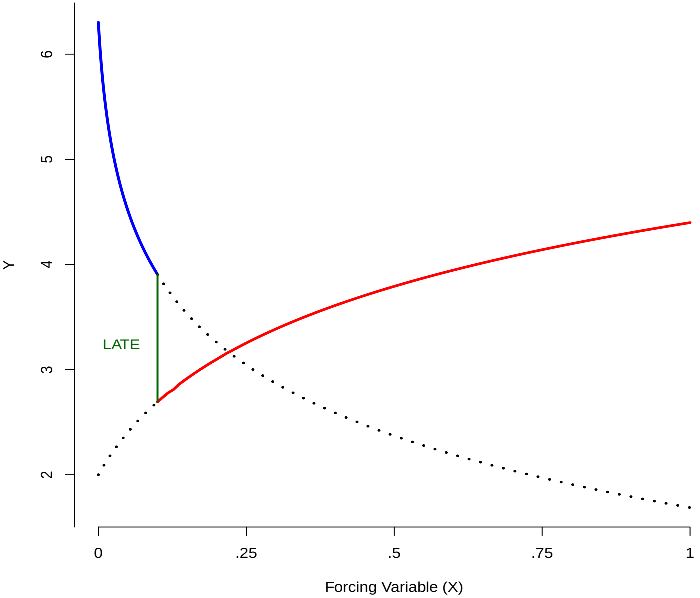
Simple Linear RD Setup



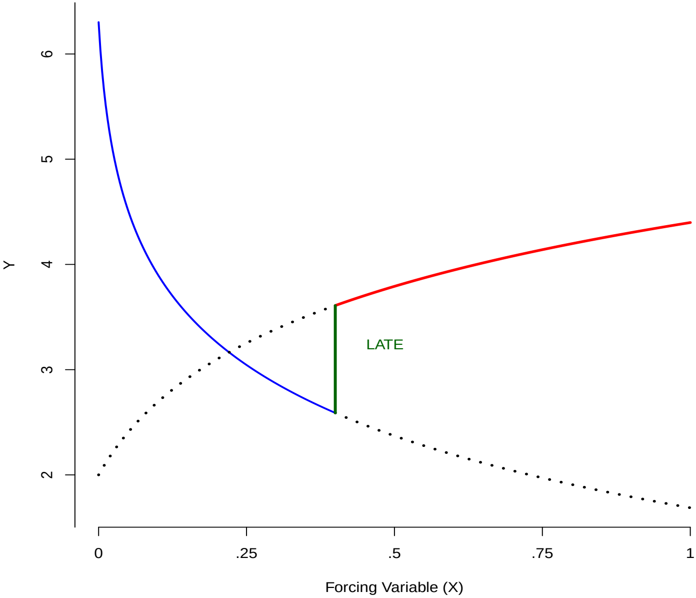
Potential Outcomes



Cutpoint Selection and LATE



Cutpoint Selection and LATE



Lee's Interpretation

Can individuals influence the assignment variable X?

$$
\begin{aligned}
Y =& T\tau + W\delta_1 + U \\
T =& I[X \geq x_0] \\
X =& W\delta_2 + V
\end{aligned}
$$

Lee's Interpretation

Can individuals influence the assignment variable X?

$$
\begin{aligned}
Y &= T\tau + W\delta_1 + U \\
T &= I[X \geq x_0] \\
X &= W\delta_2 + V
\end{aligned}
$$

- Y is the outcome, T is treatment, W is a vector of pre-treatment observables, and U are unobservables that might confound the assignment variable.

Lee's Interpretation

Can individuals influence the assignment variable X?

$$
\begin{aligned}
Y =& T\tau + W\delta_1 + U \\
T =& I[X \geq x_0] \\
X =& W\delta_2 + V
\end{aligned}
$$

- Y is the outcome, T is treatment, W is a vector of pre-treatment observables, and U are unobservables that might confound the assignment variable.
- W is endogenously determined, δ_1 and δ_2 need not be 0, no assumptions about correlations between W, U, and V.

Lee's Interpretation

- Heterogeneity in the outcome is completely described the pair of random variables (W, U)

Lee's Interpretation

- Heterogeneity in the outcome is completely described the pair of random variables $(W,\ U)$
- The distribution of X, conditional on a particular pair of values $W = w, U = u$, is equivalent (apart from an additive shift) to the distribution of V conditional on $W = w, U = u$

Lee's Interpretation

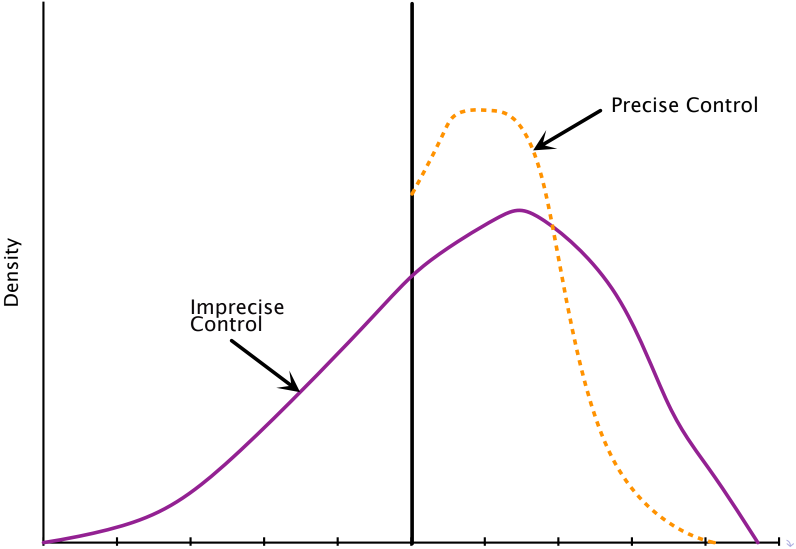
- Heterogeneity in the outcome is completely described the pair of random variables (W, U)
- The distribution of X, conditional on a particular pair of values $W = w, U = u$, is equivalent (apart from an additive shift) to the distribution of V conditional on $W = w, U = u$
- If there is some room for error, but individuals have precise control about whether they fail to receive the treatment, then the density of X will be 0 below the threshold, but positive above the threshold.

Lee's Interpretation

- Heterogeneity in the outcome is completely described the pair of random variables $(W,\ U)$
- The distribution of X, conditional on a particular pair of values $W = w, U = u$, is equivalent (apart from an additive shift) to the distribution of V conditional on $W = w, U = u$
- If there is some room for error, but individuals have precise control about whether they fail to receive the treatment, then the density of X will be 0 below the threshold, but positive above the threshold.
- If there is stochastic error in the assignment variable and individuals do *not* have precise control over the assignment variable, we would expect the density of X (and hence V), conditional on $W = w, U = u$ to be continuous at the discontinuity threshold.

Is our RD Design Valid?

Are individuals able to influence the forcing variable, and if so, what is the nature of this control?



More Lee

- **Definition:** We say individuals have imprecise control over X when conditional on $W = w$ and $U = u$, the density of V (and hence X) is continuous.

More Lee

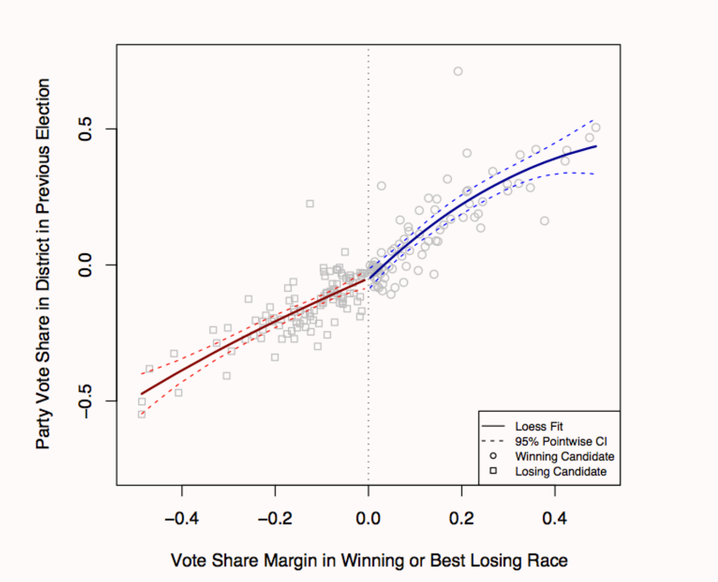
- **Definition:** We say individuals have imprecise control over X when conditional on $W = w$ and $U = u$, the density of V (and hence X) is continuous.
- **Local Randomization:** If individuals have imprecise control over X as defined above, then $P(W = w, U = u | X = x)$ is continuous in x; the treatment is "as good as" randomly assigned around the cutoff.

Eggers and Hainmueller (2009)

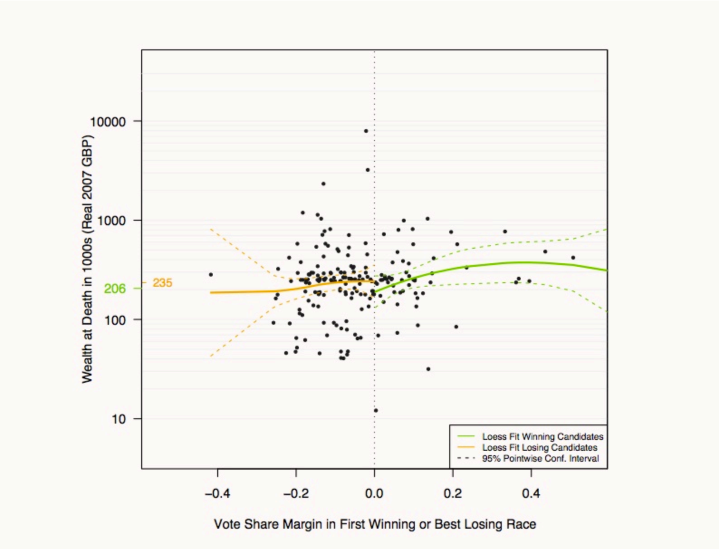
What is the personal benefit of holding office?

> *We attempt to measure this benefit by examining the effect of serving in Parliament on the estates of British politicians...(We) compare the wealth (at death) of MPs with that of politicians who ran for Parliament unsuccessfully. Voting, not randomization, decides which candidates win elections;...(We) employ a regression discontinuity design (Lee 2008; Thistlethwaite and Campbell 1960), exploiting the quasirandom assignment of office in very close races to estimate the effect of office on wealth.*

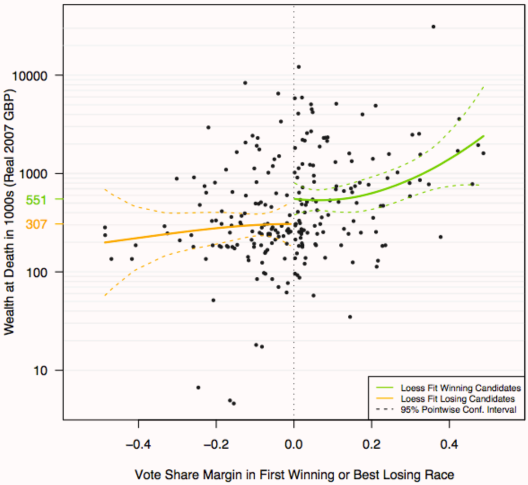
Eggers and Hainmueller (2009)



Eggers and Hainmueller: Laborites



Eggers and Hainmueller: Tories



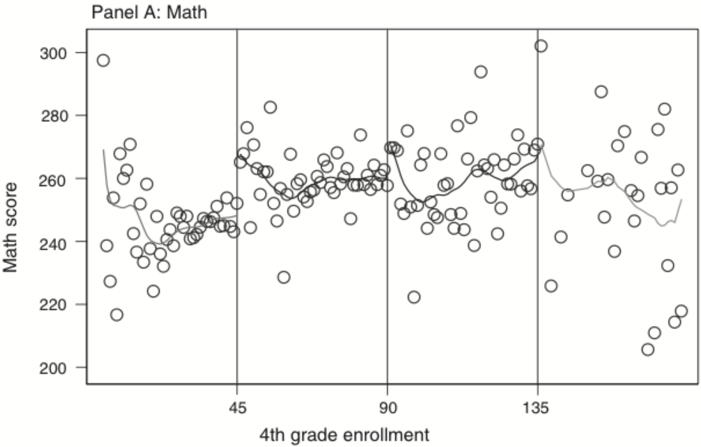
Urquiola

- Chiliean private schools cannot enroll more than 45 students per classroom.

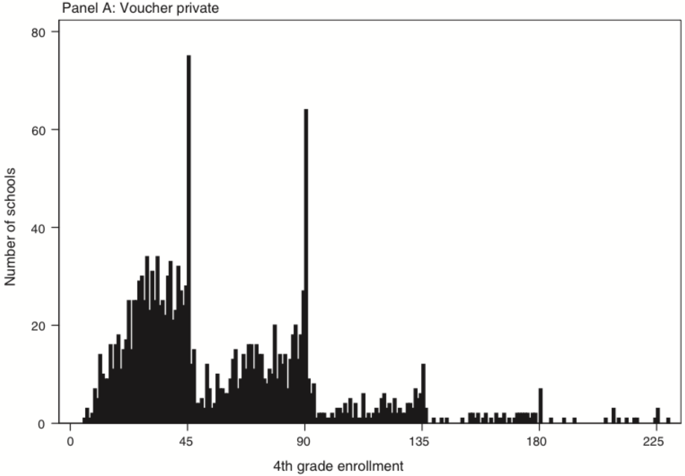
Urquiola

- Chiliean private schools cannot enroll more than 45 students per classroom.
- *. . . in the presence of the class-size cap and the integer constraint on the number of classrooms, schools at the cap adjust price (or enrollment) to avoid having an additional classroom. This results in stacking at enrollment levels that are multiples of 45. Because higher income households sort into higher-productivity schools, the stacking implies discontinuous changes in average family income and hence in other correlates of income, such as mothers' schooling, at these multiples.*

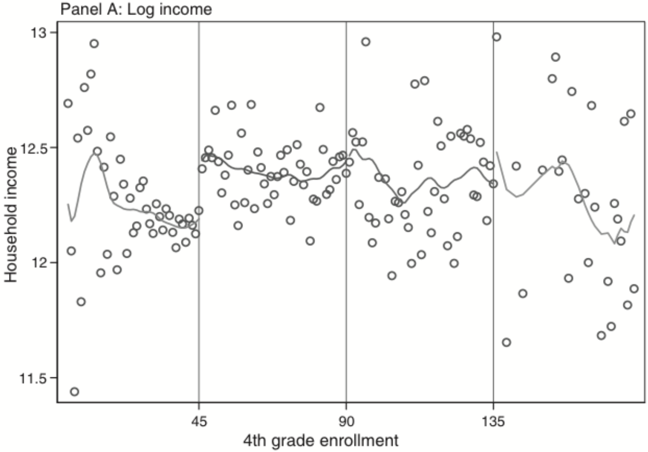
Urquiola


Panel A: Math
Math score
300
280
260
240
220
200
45
90
135
4th grade enrollment

Urquiola


Panel A: Voucher private
Number of schools
80
60
40
20
0
0
45
90
135
180
225
4th grade enrollment

Urquiola


Panel A: Log income
13
12.5
12
11.5
Household income
45
90
135
4th grade enrollment

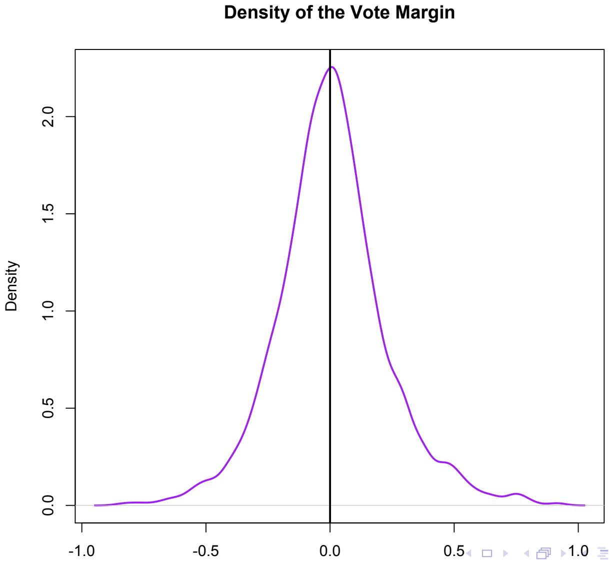
Incumbency (Dis) Advantage in Brazil

- Data used in R. Titiunik, "Incumbency Advantage in Brazil: Evidence from Municipal Mayor Elections"
- Let municipality i at election t have J political parties that dispute municipal mayor elections.
- Let V_{itj} be the vote share obtained by party j in municipality i in election t.
- The margin of victory (or loss) for party k (our forcing variable) is defined as $Z_{itk} = V_{itk} - V_{itj}$, where V_{itj} is the vote share of party k's strongest opponent.
- The rule determining incumbency status:

$$T_{it+1,k} = \begin{cases} 1 & \text{if } Z_{itk} \geq 0 \\ 0 & \text{if } Z_{itk} < 0 \end{cases}$$

Density of the Forcing Variable

Is there evidence of manipulation? No.

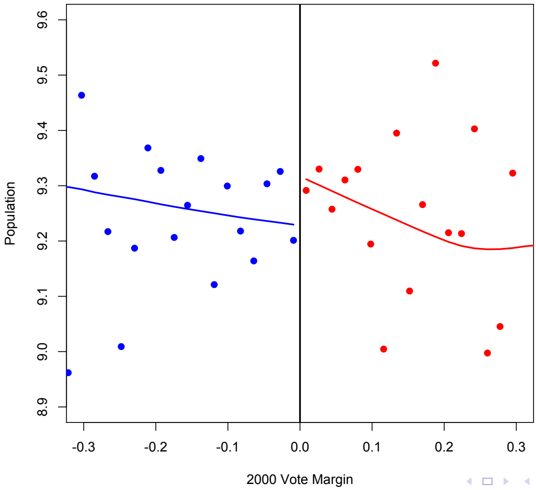


Binning

- A standard way of graphing the data is to divide the forcing variable into a number of bins and then averaging the outcome variable in each bin. The bin averages are then plotted against the bin mid-points.
- The key question is whether there is evidence of a jump in the conditional mean of the outcome at the cutoff. If there is no visual evidence of a "jump" at c, it is unlikely that more sophisticated analyses will lead to credible effect estimates that are different from 0. More formal analyses are essentially more sophisticated versions of this binning procedure.

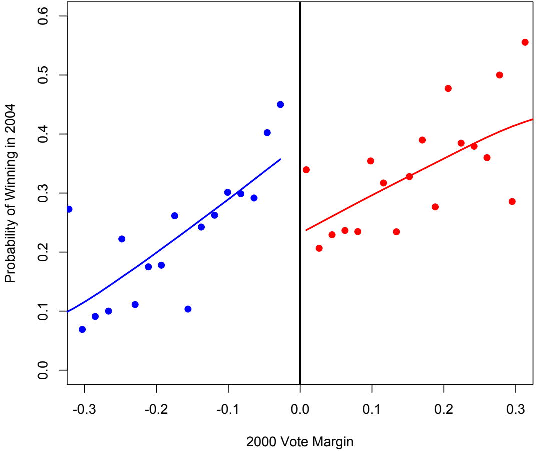
Binning

Using Pre-Treatment Covariates to check the validity of the design:



Outcome

Is there an incumbency advantage?



Kernel Estimator

- Define the conditional means: $\mu_l = E[Y(0)|X = c]$ and $\mu_r = E[Y(1)|X = c]$
- The estimand is $\tau_{rd} = \mu_r(c) - \mu_l(c)$.
- One approach is to use a kernel $K(u)$, with $\int K(u)\,du = 1$ and a bandidth of h, i.e. your "window".
- To calculate $\hat{\tau}_{RD} =$

$$\frac{\sum_{i:X_i \geq c} Y_i \cdot K((X_i - x)/h)}{\sum_{i:X_i \geq c} K((X_i - x)/h)} - \frac{\sum_{i:X_i < c} Y_i \cdot K((X_i - x)/h)}{\sum_{i:X_i < c} K((X_i - x)/h)}$$

Rectangular Kernel

- One common estimator uses a rectangular kernel, which weights each observation in the bandwidth window equally:

$$\frac{\sum_{i:X_i \geq c} Y_i \cdot 1\{c \leq X_i \leq c+h\}}{1\{c \leq X_i \leq c+h\}} - \frac{\sum_{i:X_i \geq c} Y_i \cdot 1\{c-h \leq X_i \leq c\}}{1\{c-h \leq X_i \leq c\}}$$

- This estimator can be interpreted as first throwing away observations with a value of X_i more than h away from c, and then simply differencing the average outcomes by treatment status in the remaining sample.

Rectangular Kernel

- One common estimator uses a rectangular kernel, which weights each observation in the bandwidth window equally:

$$\frac{\sum_{i:X_i \geq c} Y_i \cdot 1\{c \leq X_i \leq c+h\}}{1\{c \leq X_i \leq c+h\}} - \frac{\sum_{i:X_i \geq c} Y_i \cdot 1\{c-h \leq X_i \leq c\}}{1\{c-h \leq X_i \leq c\}}$$

- This estimator can be interpreted as first throwing away observations with a value of X_i more than h away from c, and then simply differencing the average outcomes by treatment status in the remaining sample.

Local Linear Regression

- Instead of locally fitting a constant function, we can fit linear regression functions to the observations within a distance h on either side of the discontinuity point:

$$\min \sum_{i:c-h<X_i<c} (Y_i - \alpha_l - \beta_l \cdot (X_i - c))^2$$

and

$$\min \sum_{i:c\leq X_i<c+h} (Y_i - \alpha_r - \beta_r \cdot (X_i - c))^2$$

- The value of $\mu_l(c)$ is estimated as $\hat{\mu}_l(c) = \hat{\alpha}_l + \hat{\beta}_l \cdot (c - c) = \hat{\alpha}_l$ and $\hat{\mu}_r(c)$ is estimated as $\hat{\mu}_r(c) = \hat{\alpha}_l + \hat{\beta}_l \cdot (c - c) = \hat{\alpha}_r$.
- $\hat{\tau}_{RD} = \hat{\alpha}_r - \hat{\alpha}_l$.

Cross-Validation

- How do we check the accuracy of a predictive model?
- Many predictive models tend to over-fit, so good practice is to choose a predictive model based on a training data set and then check its predictive accuracy on a seperate validation dataset.
- Training datasets aren't available, as in our regression discontinuity case, but one useful technique for testing our model's predictive accuracy is known as **cross-validation**.

Cross-Validation

- We usually refer to prediction error as the expected squared difference between a future response and its prediction from the model:

$$\mathrm{PE} = E\{(y - \hat{y})^2\}$$

- In cross validation, we use part of the data to fit the model, and a different part to test it.
- Suppose we split the data into K parts. Let $k(i)$ be the part containing observation i. Denote the by $\hat{y}_i^{-k(i)}$, the fitted value for observation i, computed with the $k(i)$th part of the data removed. Then the cross-validation prediction error is:

$$\mathrm{CV} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i^{-k(i)})^2$$

"Leave-One-Out"

- Often we choose $k = n$, resulting in **"leave-one-out"** cross-validation.
- For each observation i, we refit the model leaving that observation out of the data, and then compute the predicted value for the ith observation and compute the predicted value $\hat{y}_i^{-i}$. We do this for each observation and then compute the average cross-validation sum of squares $\mathrm{CV} = \sum(y_i - \hat{y}_i^{-i})^2/n$

Regression Discontinuity

Basic Setup

Lee's Interpretation

Examples

Nuts and Bolts

Graphical Methods

Analysis

Cross-Validation

What are we predicting?

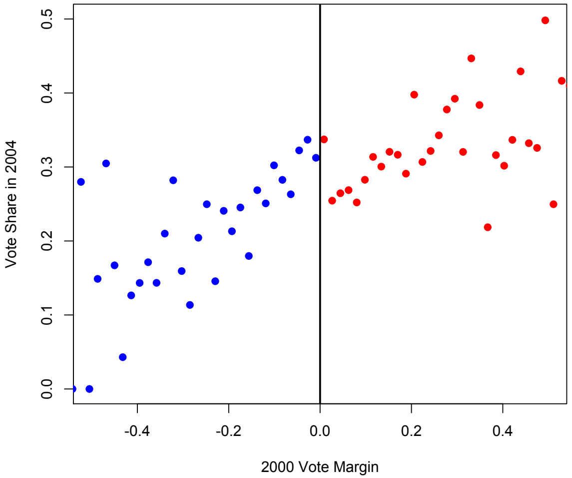
$$\min \sum_{i:c-h<X_i<c} (Y_i - \alpha_l - \beta_l \cdot (X_i - c))^2$$

and

$$\min \sum_{i:c \le X_i < c+h} (Y_i - \alpha_r - \beta_r \cdot (X_i - c))^2$$

- The value of $\mu_l(c)$ is estimated as $\hat{\mu}_l(c) = \hat{\alpha}_l + \hat{\beta}_l \cdot (c - c) = \hat{\alpha}_l$ and $\hat{\mu}_r(c)$ is estimated as $\hat{\mu}_r(c) = \hat{\alpha}_r + \hat{\beta}_r \cdot (c - c) = \hat{\alpha}_r$.
- $\hat{\tau}_{RD} = \hat{\alpha}_r - \hat{\alpha}_l$.

Effect of Incumbency on Vote Share



Picking h

- We need to pick an h and cross-validation is a natural "hands-off" technique.
- Predict each y_i using x_i values within h. Note that we each treat each y_i as point at a boundary.
- To emulate the fact that RD estimates are based on regression estimates at the boundary, the regression is estimated using only observations with values of X on the left of X_i ($X_i - h \leq X < X_i$) for observations on the left of the cutpoint ($X_i < c$). For observations on the right of the cutoff point ($X_i \geq c$), the regression is estimated using only the observations with values of X on the right of X_i ($X_i < X \leq X_i + h$)

- Formally, let $\hat{Y}(X_i)$ be the predicted value of Y obtained using the regressions described above. The cross validation criterion is defined as

$$\mathrm{CV}_Y(h) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}(X_i))^2$$

with the corresponding cross-validation choice for the bandwidth

$$h_{\mathrm{CV}}^{\mathrm{opt}} = \arg\min_h \mathrm{CV}_Y(h)$$

Results of Cross-Validation