Regression Discontinuity

October 7, 2009

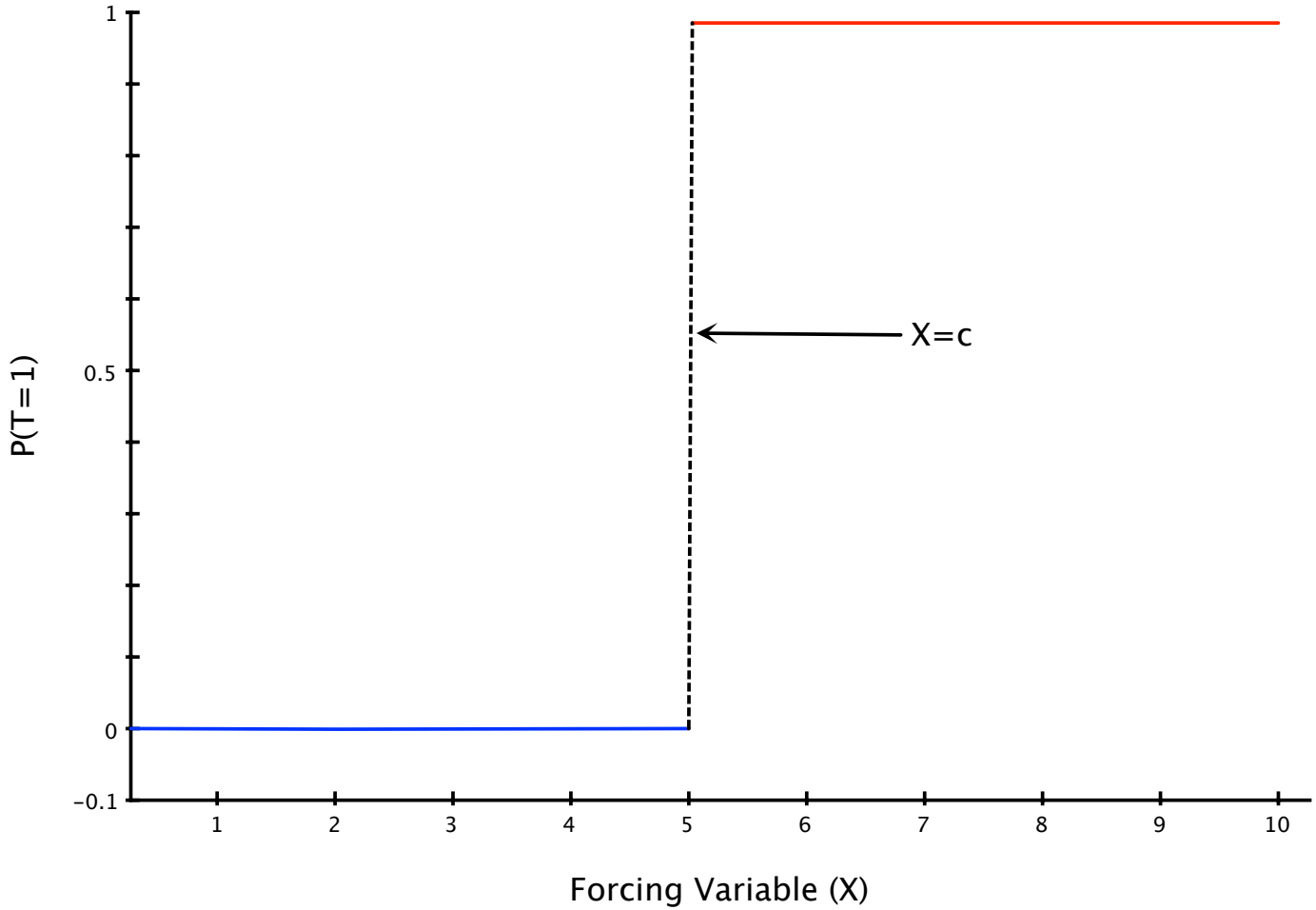
Living on the Edge

- $T_i = 1[X_i \geq c]$
- $\tau_{rd} = E[Y_i(1) - Y_i(0)|X_i = c]$

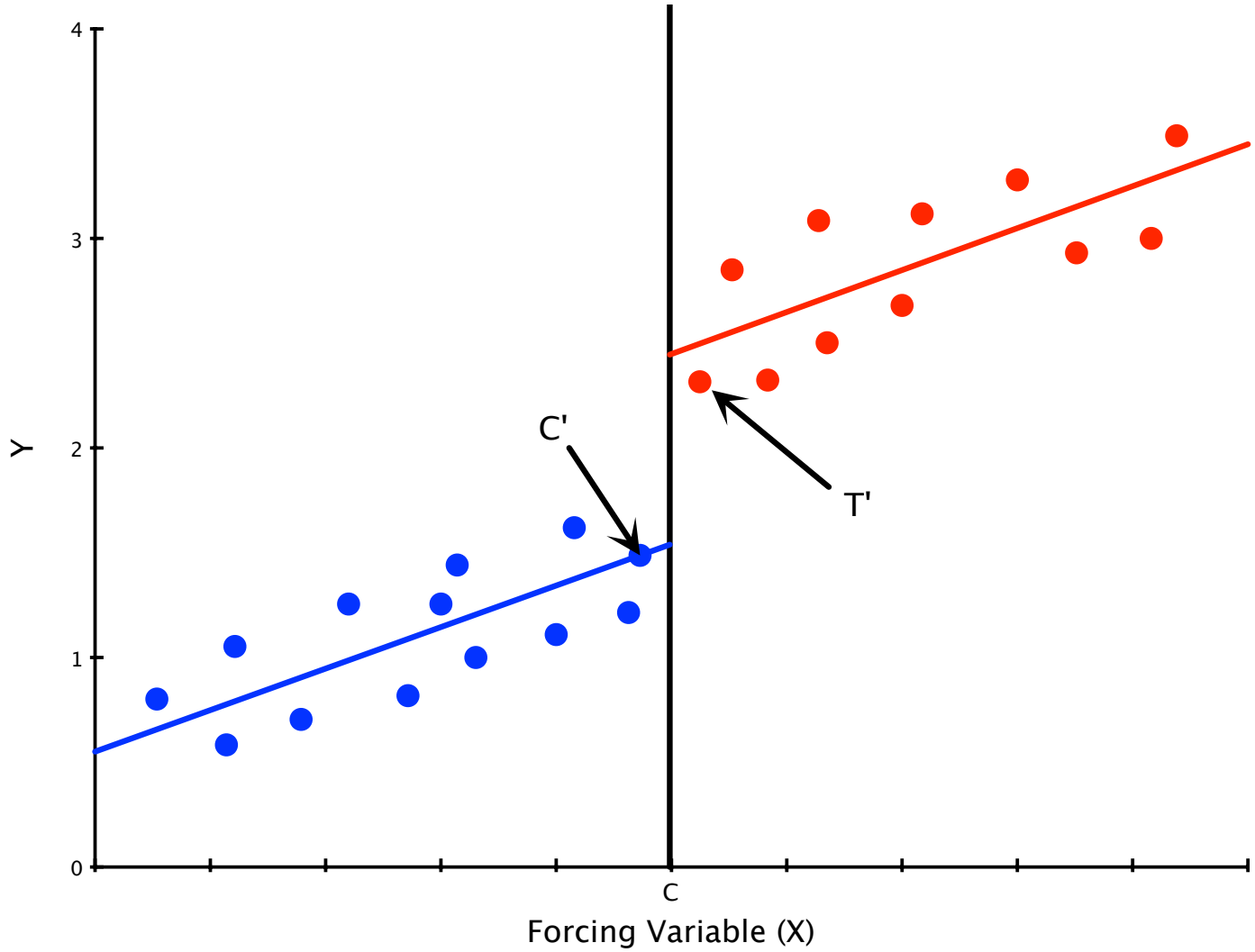
Assumptions

1. $Y_i(0), Y_i(1) \perp\!\!\!\perp T_i|X_i$
   - 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

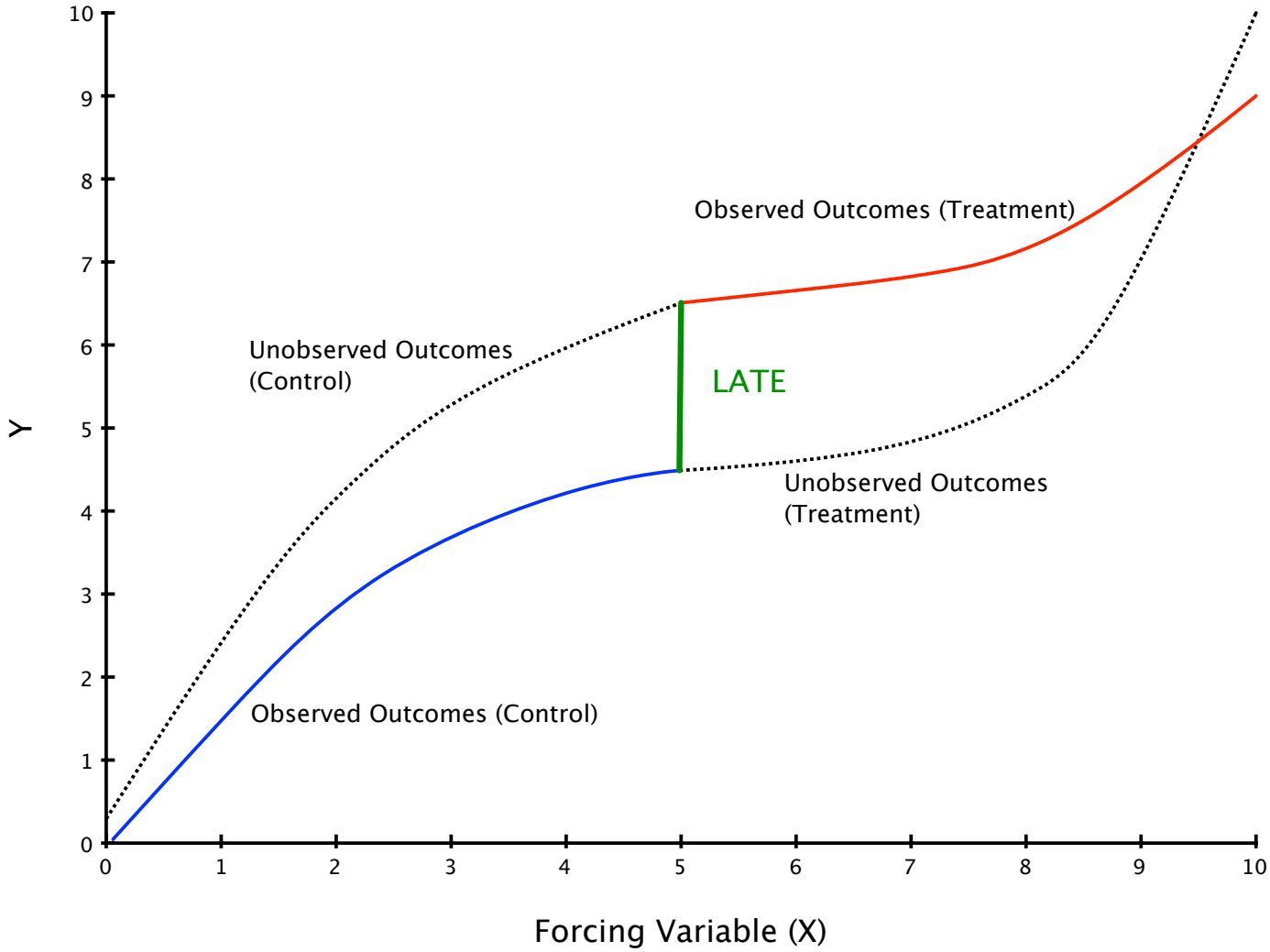
Probability of Treatment



Simple Linear RD Setup



Potential Outcomes



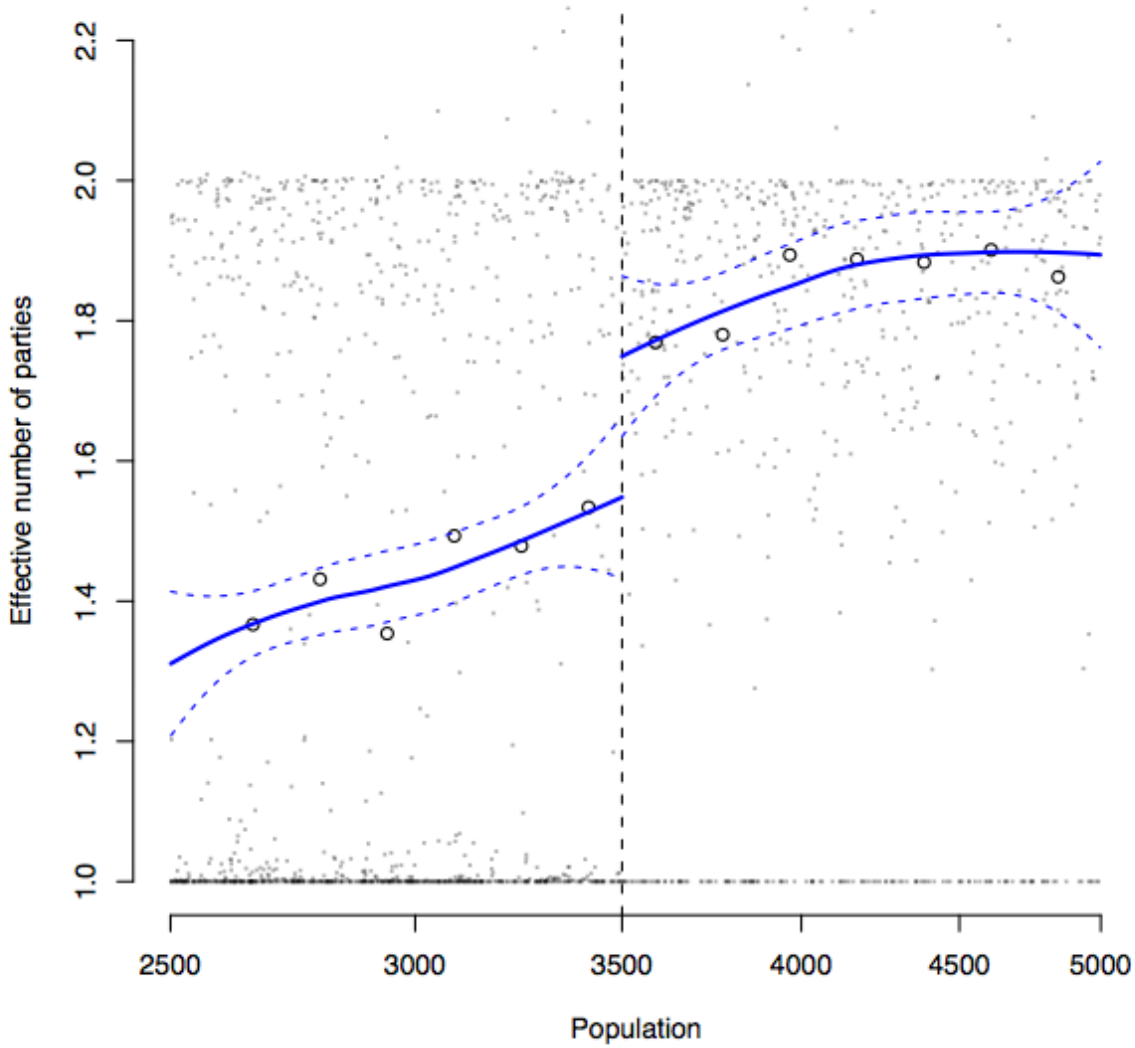
Eggers (2009)

Duverger's Law (1972):

> *To these socio-economic and historical factors a technical factor must be added: the electoral system. I expressed its effects in 1946 in the formulation of three sociological laws: (1) a majority vote on one ballot is conducive to a two-party system; (2) proportional representation is conducive to a multiparty system; (3) a majority vote on two ballots is conducive to a multiparty system, inclined toward forming coalitions.*

In French municipalities, the electoral rule used to elect the municipal council depends on the city's population: cities with fewer than 3,500 people elect their councils by a form of plurality rule, while those with a population of 3,500 or more use a form of PR.

Eggers (2009)


Effective number of parties
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2500
3000
3500
4000
4500
5000
Population

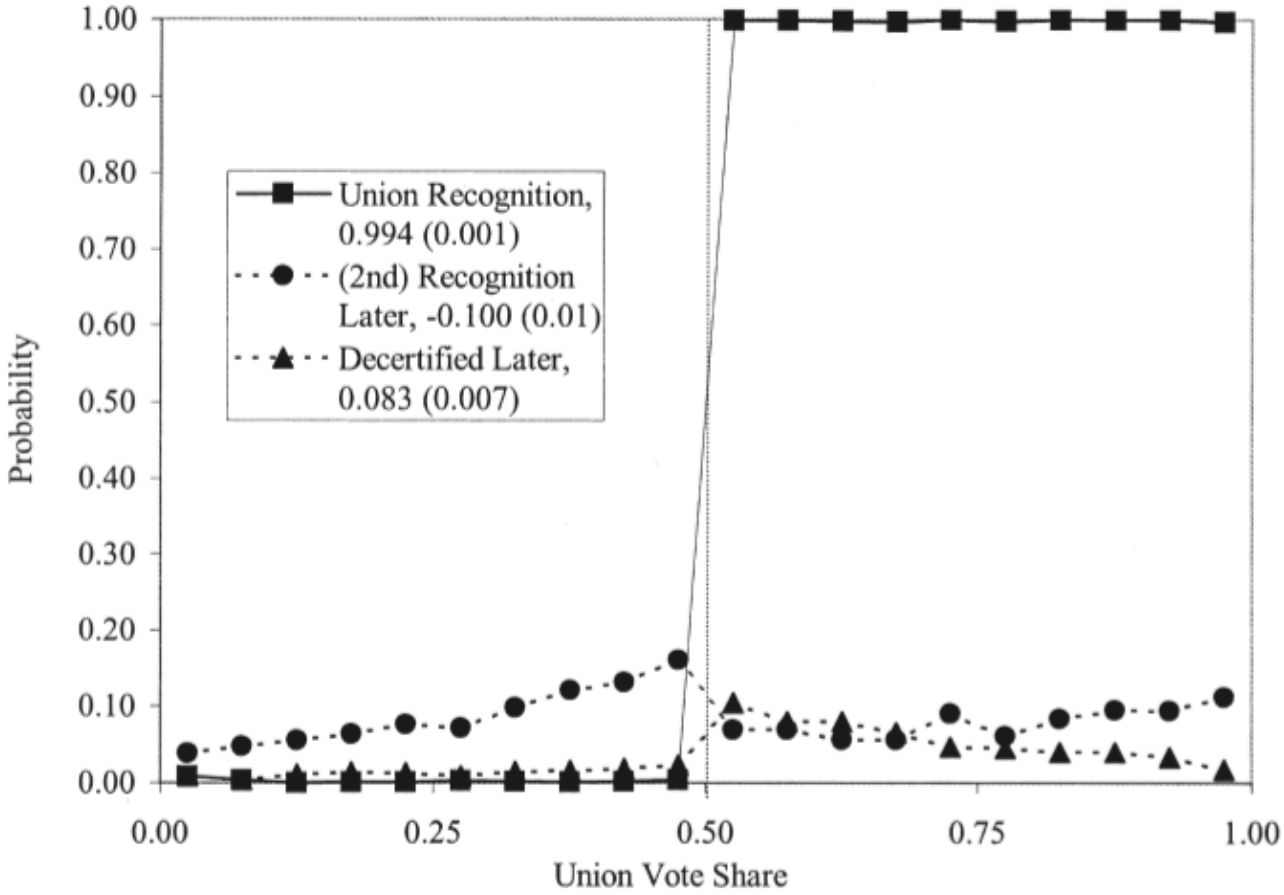
Lee and Dinardo (2004)

Lee and Dinardo document that the outcome of an NLRB election has a substantial, binding impact on the collective bargaining pro- cess, even among close elections. Where they barely win the election, unions are able to maintain their legal recognition over long time horizons; where they barely lose, there is little evidence of subsequent attempts to organize the workplace.

Question: **What is the impact of union recognition keeping all other things—including having held an elections—equal?**

Lee and Dinardo (2004)

Do NLRB elections lead to unionization?



Does unionization increase wages?

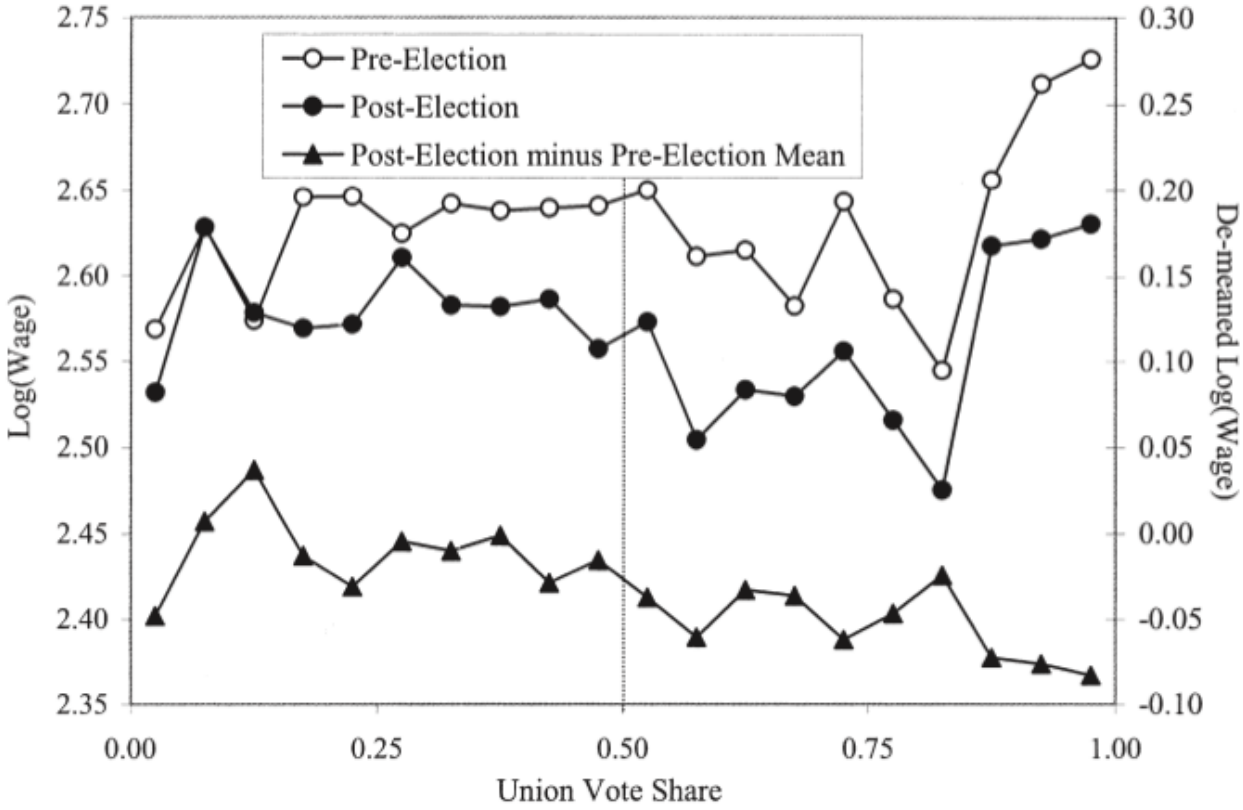
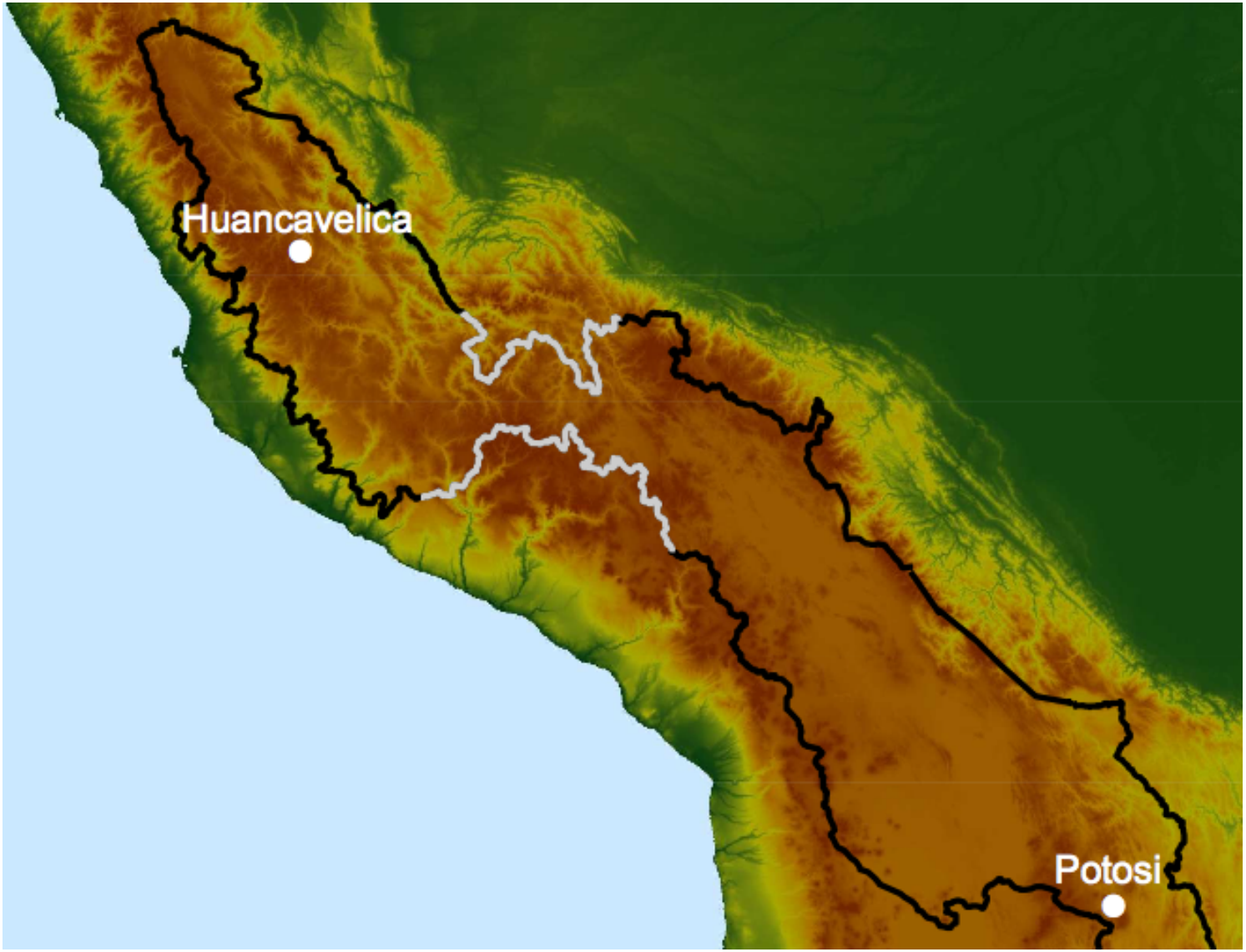


FIGURE IXb
Log(Production Hourly Wage), Pre- and Postelection, by Union Vote Share, LRD

Dell (2009)

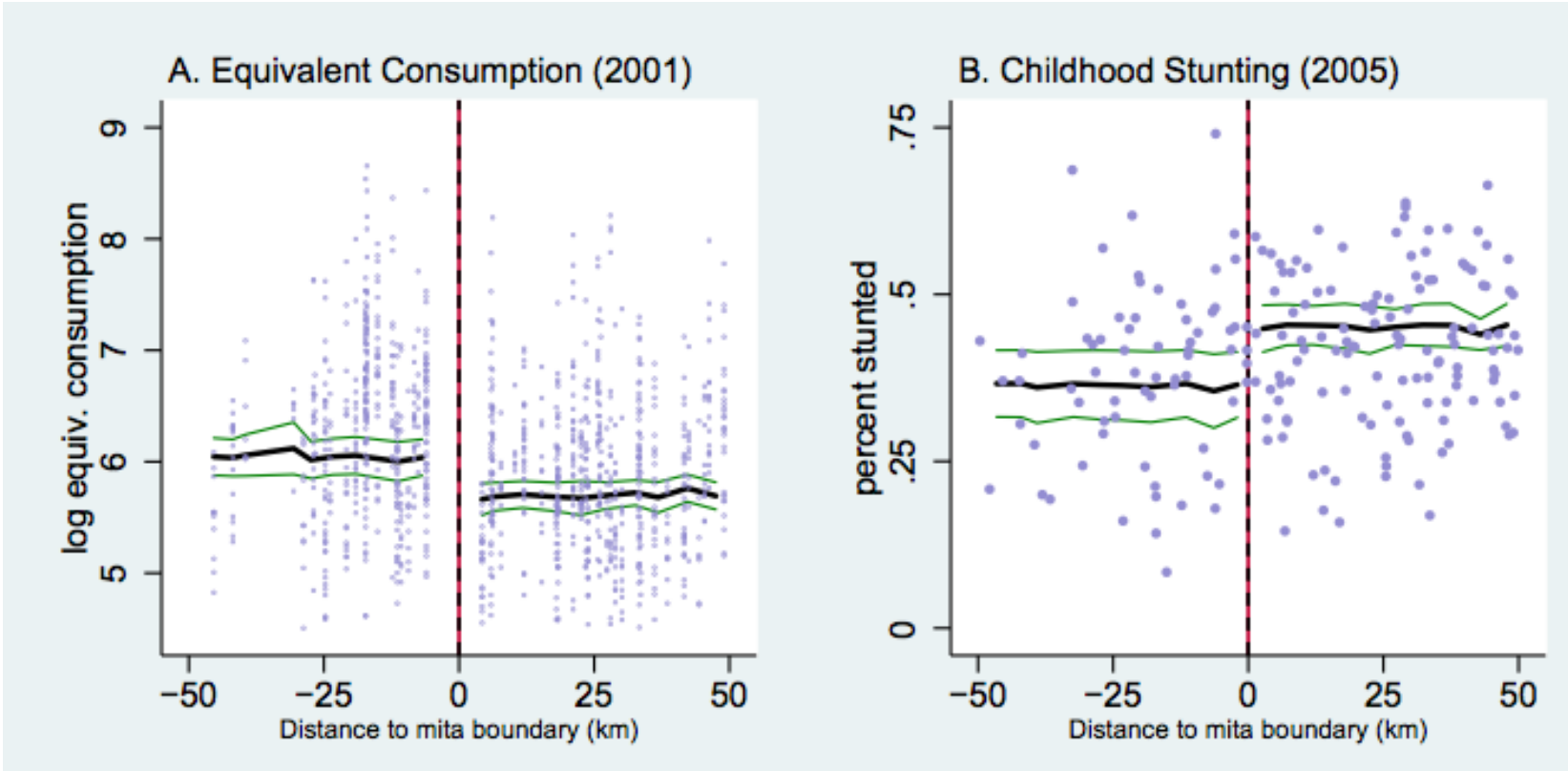
- What are the long term impacts of colonial institutions?
- Dell (2009) examines the long run impacts of the mining mita, a forced labor system instituted by the Spanish government in Peru and Bolivia in 1573 and abolished in 1812. The mita required over 200 indigenous communities to send one seventh of their adult male population to work in the Potosi silver and Huancavelica mercury mines
- *The contribution of mita conscripts changed discretely at the boundary of the subjected region - on one side all communities sent the same percentage of their population to the mines, while on the other side all communities were exempt*

Dell (2009)



Dell (2009)

The long term impact of the Mita on development outcomes:



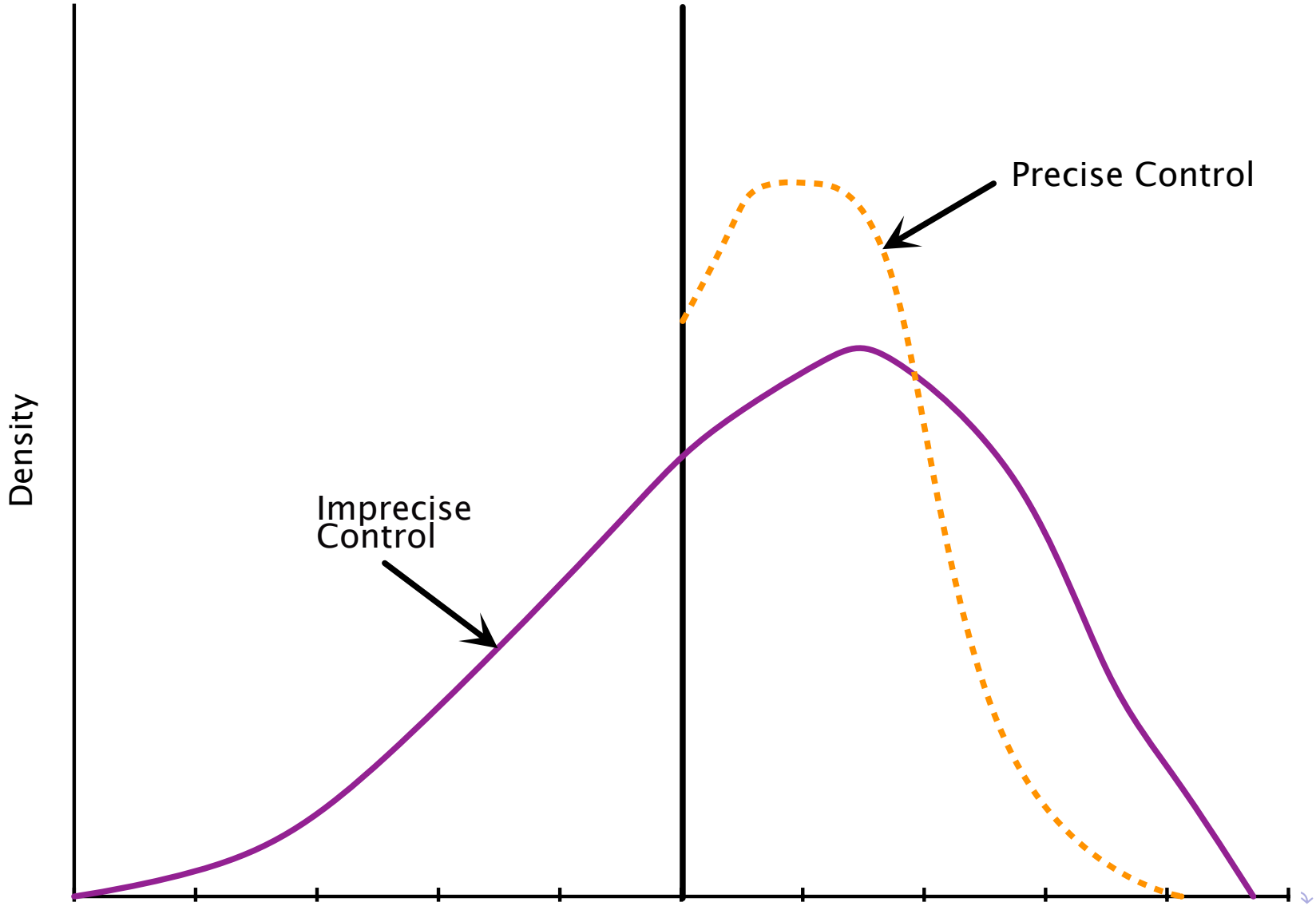
Incumbency (Dis) Advantage in Brazil

- 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:

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

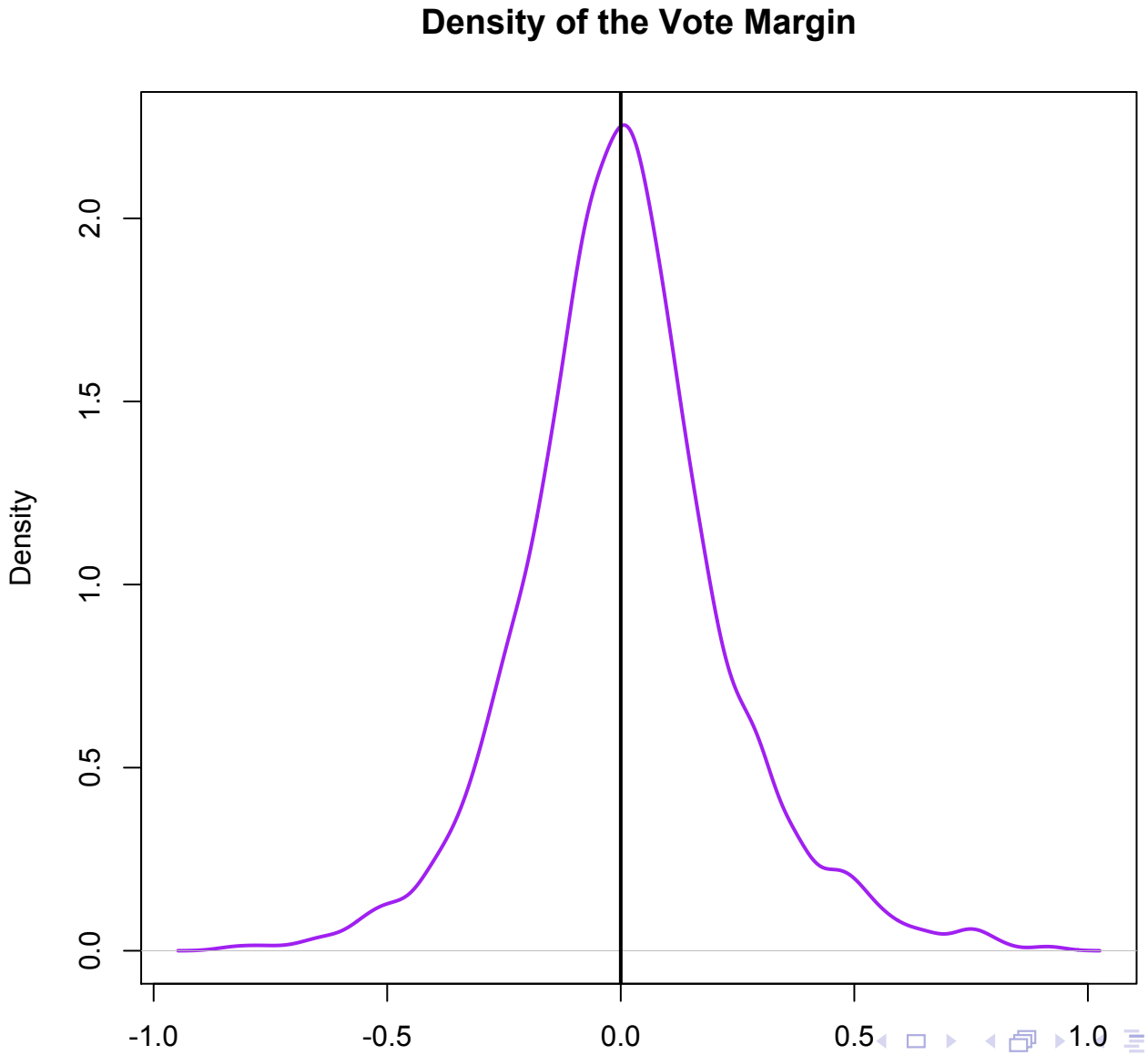
Is our RD Design Valid?

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



Density of the Forcing Variable

Is there evidence of manipulation? No.

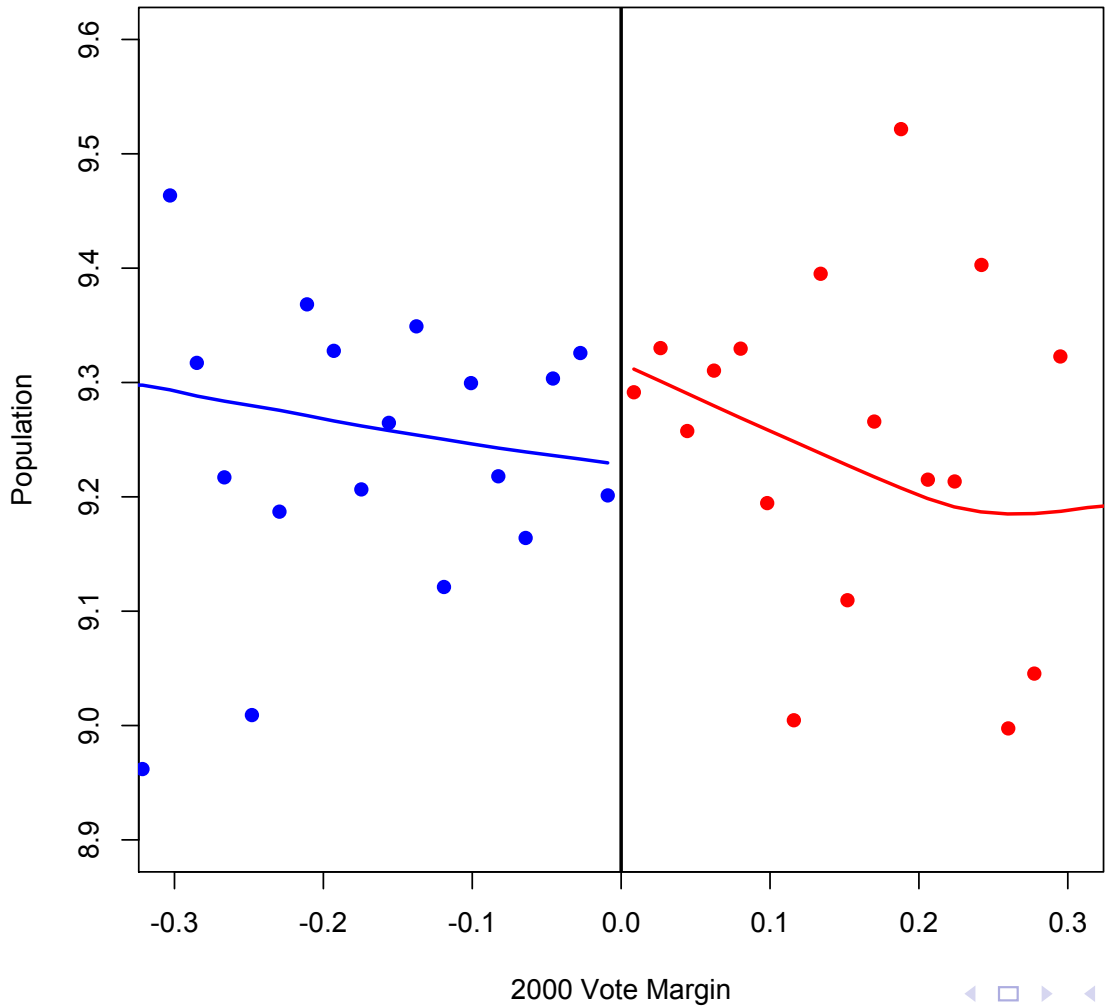


Binning

- 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.

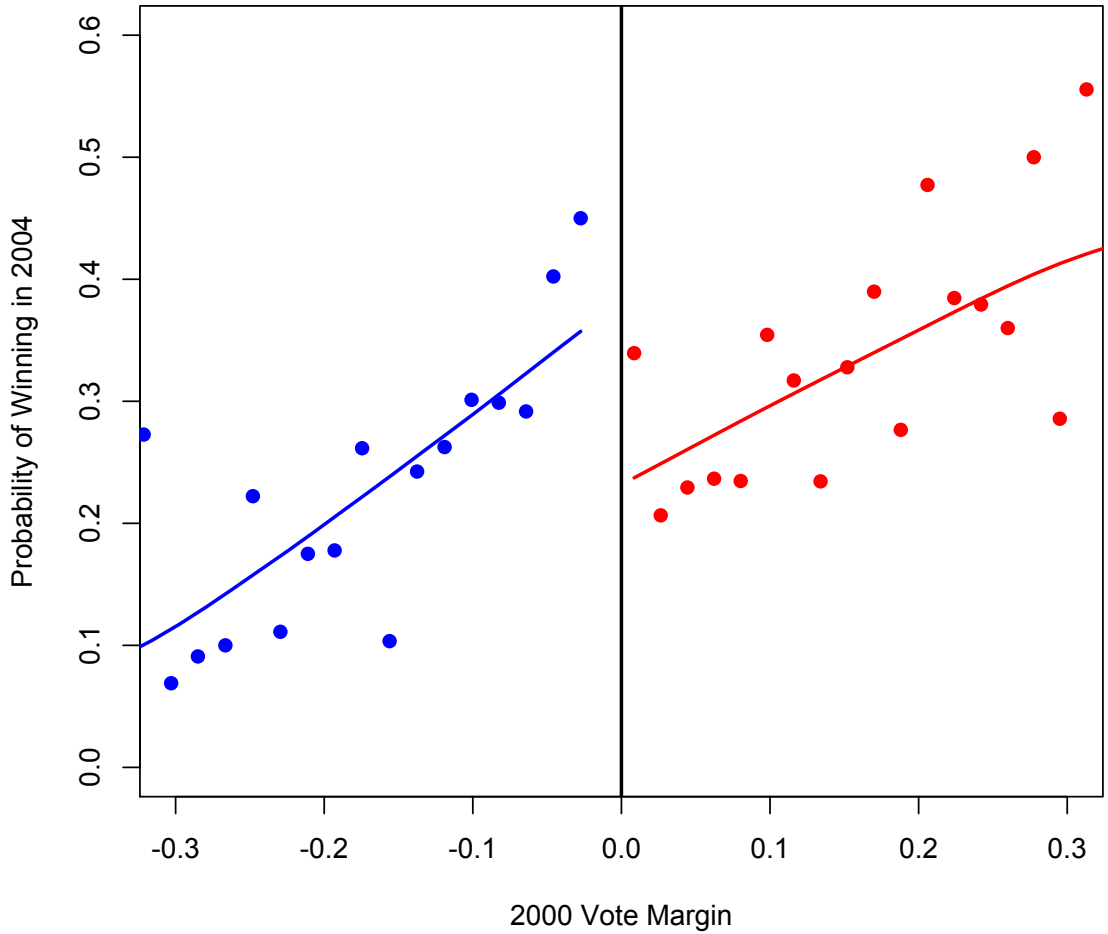
Binning

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 $\hat{\mu}_r(c) = \hat{\alpha}_l + \hat{\beta}_l \cdot (c - c) = \hat{\alpha}_r$.
- $\hat{\tau}_{RD} = \hat{\alpha}_r - \hat{\alpha}_l$.