Difference-in-differences & Regression Discontinuities

Felipe Balcazar

NYU

May , 2023

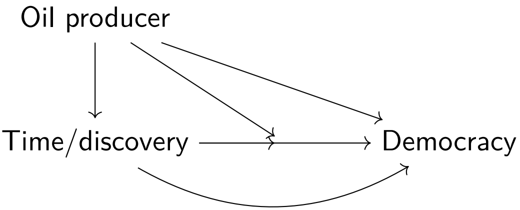
NEW YORK UNIVERSITY

Using time to identify the causal effect

- The time dimension is useful, we can control for time-invariant confounding. But can do more with it?
- In our example, some individuals get a masters' degree while some others don't.
- The treatment interacts with time!

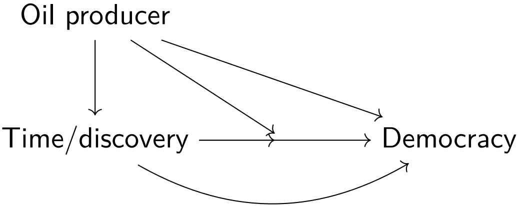
Using time to identify the causal effect

- The time dimension is useful, we can control for time-invariant confounding. But can do more with it?
- In our example, some individuals get a masters' degree while some others don't.
- The treatment interacts with time!



Using time to identify the causal effect

- The time dimension is useful, we can control for time-invariant confounding. But can do more with it?
- In our example, some individuals get a masters' degree while some others don't.
- The treatment interacts with time!



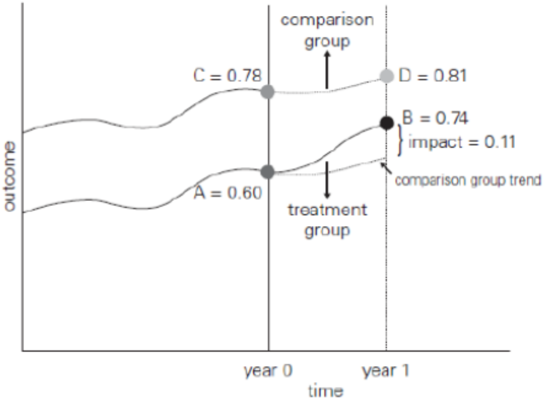
Can we use this to identify causal effects?

$$Dem_{it} = \beta_1 Oil_i + \beta_2 disc_{it} + \beta_3 (oil_i \times disc_{it}) + \gamma_{it} + \varepsilon_{it}$$

Difference-in-Differences Estimates

- It is one particular form of fixed effects, within-unit analysis that is very easy to interpret.
- Applicable whenever you have one policy/intervention affecting a treated group (affected by the change) but not a control group (not affected by the change) at one specific point in time.
- Assumption: *changes* in outcomes before and after policy change for control group, provide counterfactual for what would have been the change for the treated, had they not been treated.
- No need to assume that treated and control group are similar on average (in levels), only that *changes* over time in their outcomes are comparable (parallel trends assumption).

Introducing differences-in-differences (DID)



- 1st difference: Post-treatment gap between groups.
- 2nd difference: Pre-treatment gap between groups.
- Difference-in-differences:

$$(\text{2nd difference}) - (\text{1st difference})$$

A simple example

$$y_{it} = c + \beta_1 Treated_i + \beta_2 Post_t + \beta_3 (Treated_i \times Post_t) + \varepsilon_{it}$$

y_{it}	$Treated = 1$	$Treated = 0$	Difference
$Post_t = 1$	y_{22}	y_{12}	$y_{22} - y_{12}$
$Post_t = 0$	y_{21}	y_{11}	$y_{21} - y_{11}$
Change	$y_{21} - y_{22}$	$y_{11} - y_{12}$	$(y_{11} - y_{21}) - (y_{12} - y_{22})$

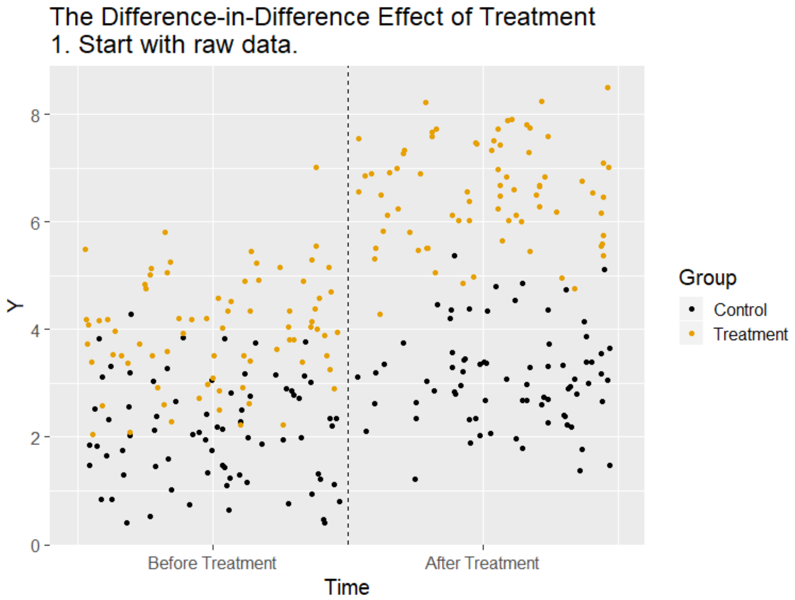
A simple example

$$y_{it} = c + \beta_1 Treated_i + \beta_2 Post_t + \beta_3 (Treated_i \times Post_t) + \varepsilon_{it}$$

y_{it}	$Treated = 1$	$Treated = 0$	Difference
$Post_t = 1$	y_{22}	y_{12}	$y_{22} - y_{12}$
$Post_t = 0$	y_{21}	y_{11}	$y_{21} - y_{11}$
Change	$y_{21} - y_{22}$	$y_{11} - y_{12}$	$(y_{11} - y_{21}) - (y_{12} - y_{22})$

This only works if the parallel trends assiumption holds!

What does DID do?



DID: Specifics

- The DID estimate is given by coefficient β_3 in the following regression:

$$Y_{it} = \alpha + \beta_1 Treated_i + \beta_2 Post_t + \beta_3 Treated_i \times Post_t + \epsilon_{it}$$

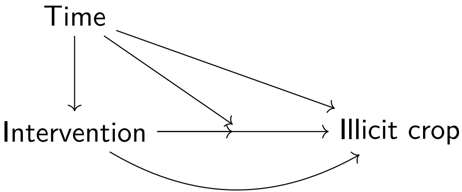
or alternatively:

$$Y_{it} = \alpha + \beta_3 Treated_i \times Post_t + \gamma_i + \rho_t + \epsilon_{it}$$

- Limitations:
  - Relies on parallel trends assumptions that cannot always be verified.
  - Picking the right control group is not always straightforward.

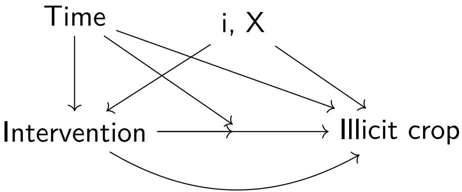
Limitations of DID

- It rests heavily on the parallel trends assumption.
- Controlling for fixed-effects helps!
- Controlling for time-varying confounders helps.



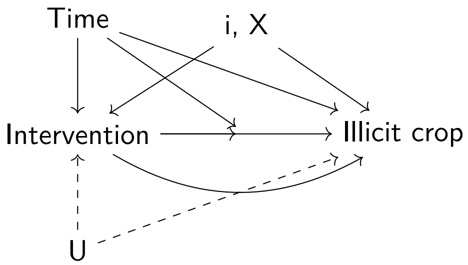
Limitations of DID

- It rests heavily on the parallel trends assumption.
- Controlling for fixed-effects helps!
- Controlling for time-varying confounders helps.



Limitations of DID

- It rests heavily on the parallel trends assumption.
- Controlling for fixed-effects helps!
- Controlling for time-varying confounders helps.



- The problem is whether to meet the parallel trends assumption we would need to control for time-varying unobservable confounders.

A world of thresholds

- You can jump in a roller-coaster when you're above 48 in.
- When you turn 21, you're legally allowed to drink.
- You need certain SAT scores to be eligible for scholarships.
- Eligibility for welfare programs based on income thresholds.
- Candidates are selected into office when they get more than half of all votes.

Thresholds as valuable experiments

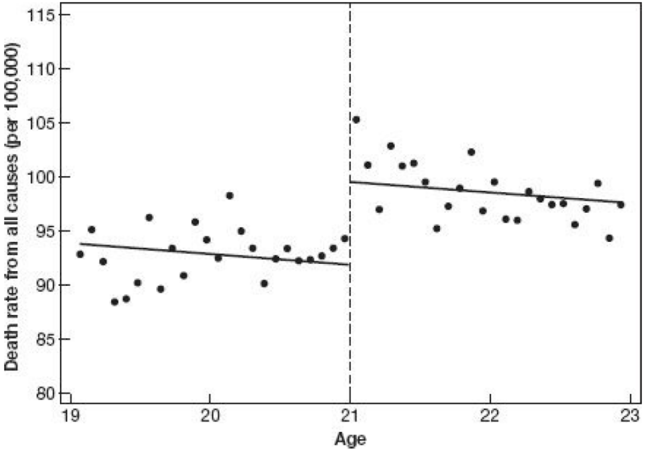
- Treatment status often varies depending on position relative to a threshold.
- Treatment status is a discontinuous function of the threshold.
  - When treatment status fully depends on position relative to threshold: sharp RD design.
  - When position relative to treatment affects *likelihood* of treatment: fuzzy RD design.
- Arbitrary rules or assignment based on thresholds can become valuable experiments.
- The assumption is that *around* the threshold all else is equal!

RD in DAG

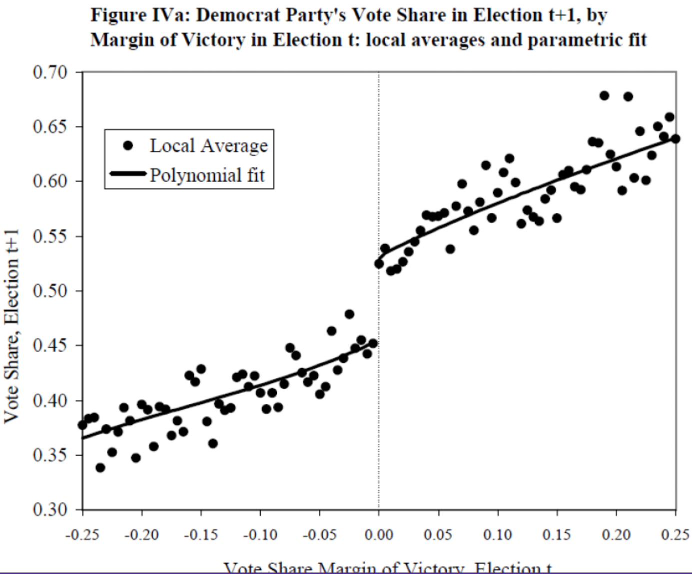
In DAG form:

$$\begin{tikzcd} D \arrow[r] & Y \\ R_i > c \arrow[u] \arrow[ur] & \end{tikzcd}$$

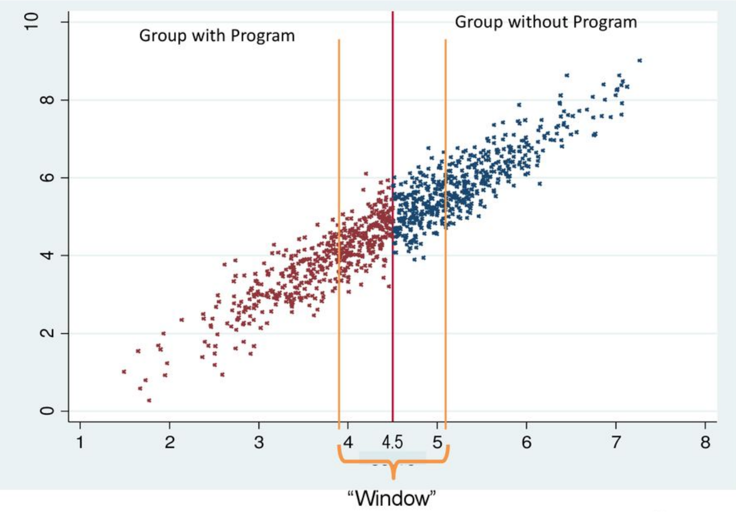
Example: Mortality rates



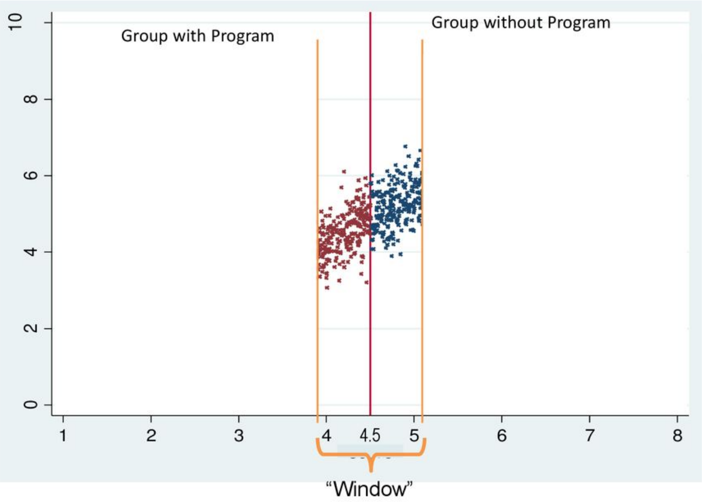
Example: incumbency advantage



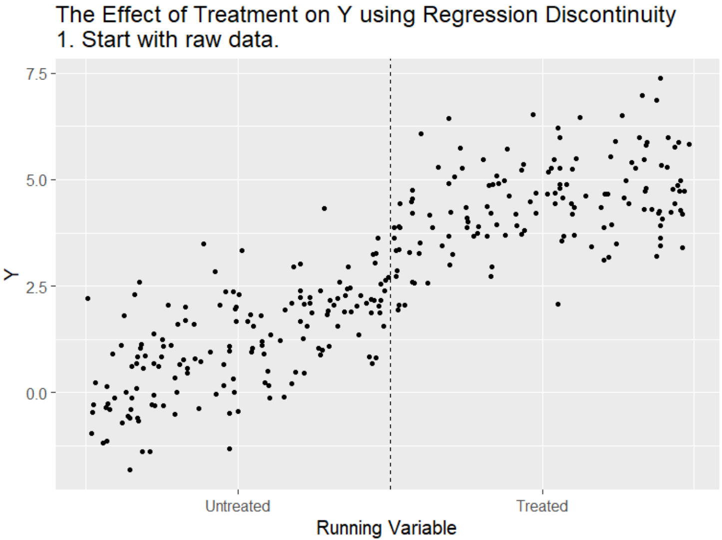
Comparisons may need to be restricted about a window



Comparisons may need to be restricted about a window



What does RDD do?

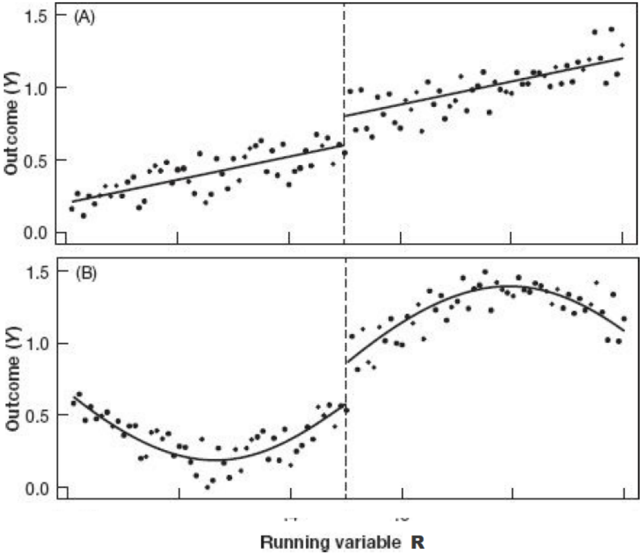


The RDD regression equation

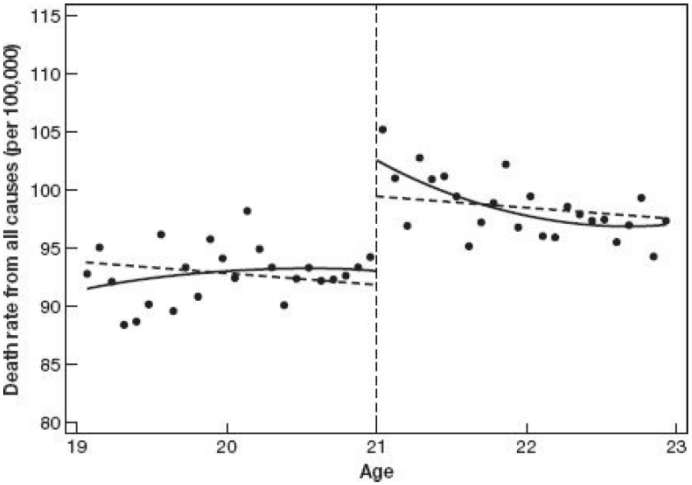
$$Y_i = \alpha + \beta D_i + \delta_1 f(R_i - c) + \delta_2 f(c - R_i) + \epsilon_i$$

- c is the threshold value.
- R_i is called the running variable.
- $D_i = 1$ if $R_i \geq c$, $D_i = 0$ otherwise.
- $f(\cdot)$ is some polynomial function.

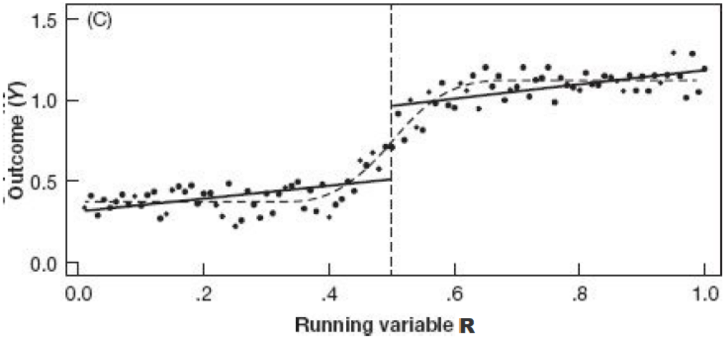
RDD and Polynomial



The functional form is important I



The functional form is important II



Limitations of RDD

- You have to worry about sorting!
    - No discontinuities for confounders.
    - This is similar to a balance test.
- The functional form must be correctly modeled.
    - Add flexibility with polynomials.
- Must be certain there are no other coinciding treatments.
- Choice of window (bandwidth) is important.

Using thresholds

- Treatment assignment may depend on meeting a requirement.
- Let's think about incumbency advantage.
- Being an incumbent implies winning an election.
- Being an incumbent provides resources.

$$\begin{tikzcd}
\text{Incumbent} \arrow[r] & \text{Re-election} \\
1st\ election > 50\% vote \arrow[u] \arrow[ur] &
\end{tikzcd}$$

Using thresholds

- Treatment assignment may depend on meeting a requirement.
- Let's think about incumbency advantage.
- Being an incumbent implies winning an election.
- Being an incumbent provides resources.

$$
\begin{tikzcd}
\text{Incumbent} \arrow[r] & \text{Re-election} \\
1st\ election > 50\% vote \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

In regression form:

$$Re - elected_i = \alpha + \beta incumbent_i + f(vote_i) + \epsilon_i$$

Using thresholds

- Treatment assignment may depend on meeting a requirement.
- Let's think about incumbency advantage.
- Being an incumbent implies winning an election.
- Being an incumbent provides resources.

$$
\begin{tikzcd}
\text{Incumbent} \arrow[r] & \text{Re-election} \\
1st\ election > 50\% vote \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

In regression form:

$$Re - elected_i = \alpha + \beta incumbent_i + f(vote_i) + \epsilon_i$$

- Is this a sharp or fuzzy RDD?

Confounding may still be an issue

- A discontinuity is a natural experiment of sorts.
- So there can be still be confounders.
- These confounders can generate *sorting*.

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
\text{Incumbent} \arrow[r] & \text{Re-election} \\
1st\ election > 50\% vote \arrow[u] \arrow[ur] &
\end{tikzcd}
$$

So we control for it:

$$Re - elected_i = \alpha + \beta incumbent_i + \delta X_i + f(vote_i) + \epsilon_i$$

Geographic RDs

- Geographic boundaries are useful thresholds (what is the running variable?).
- Across geographic boundaries you may have changes in institutions, laws, regulations (often have a bundled treatment).
- Nonetheless boundaries often help isolate the role of institutions, policies, culture, etc.

Long Run Effects of the Mita: Dell (2010)

- Dell (2010) studies the long-run effect of the Mita in Peru by comparing development outcomes in villages within the catchment area to villages barely outside the catchment area.
- Mita: required over 200 indigenous communities within the catchment area to send one-seventh of their adult male population to work in the Potosi silver and Huancavelica mercury mines.
- Focus on border segment on which RD assumptions are plausibly random.

Mita Boundaries

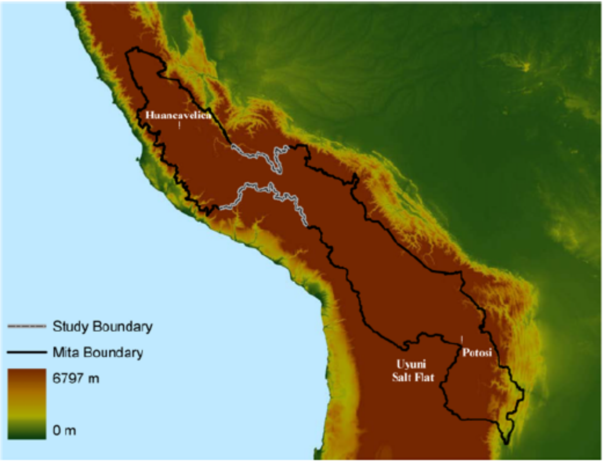


FIGURE 1.—The *mita* boundary is in black and the study boundary in light gray. Districts falling inside the contiguous area formed by the *mita* boundary contributed to the *mita*. Elevation is shown in the background.

Results

- Mita effect lowers household consumption by around 25% and increases the prevalence of stunted growth in children by around 6 percentage points.
- Fewer commercial agriculture (residents are substantially more likely to be subsistence farmers).
- Lower educational attainment.
- Less integrated into road networks.

RDs and Close Elections

- Used for understanding the causal effect of the identity of incumbents (role of gender, party, etc) on policies and other economic outcomes.
- Key assumptions is that the outcome of close races (those decided by narrow margins) is as good as random.
- Municipalities in which a given type of candidate narrowly wins or narrowly loses an election are otherwise equal.
- Can also use them to estimate the causal effect of winning and serving office on several individual outcomes.

Querubin and Snyder (2013)

- Major role of political institutions: the control of politicians.
- Major empirical question: understanding the **environments** or conditions under which democratic political institutions may be **less effective** at controlling politicians.

Rent-seeking

- One possible way to study the degree to which political institutions succeed at keeping politicians accountable is to establish the extent of systematic rent-seeking.
- The study of rent extraction faces substantial empirical challenges: it is often difficult to detect or measure the accumulation of rents by politicians in a systematic way.
- One way to assess the magnitude of political rents is to track the wealth of politicians.

Selection into Politics

The main problem underlying the estimation of the rents of a seat in Congress is selection into politics:

- More talented individuals may find holding office more costly: naive comparison *underestimates* the rents from holding office.
- More talented individuals (who would have been very successful in the private sector anyway), manage to win elections: naive comparison *overestimate* the rents from holding office.

Regression Discontinuity Design

To estimate a causal effect of political office-holding on wealth accumulation we employ an regression discontinuity design and estimate regressions of the form:

$$Wealth_i^t = \beta_0 + \beta_1 Wealth_i^{t-10} + \beta_2 Winner_i^t + \beta_3' \mathbf{X}_i + f(VoteShare_i) + \epsilon_i^t$$

$$\text{for all } i \text{ such that } |0.5 - VoteShare_i| < h$$

RDD Specifics

- We focus on the candidate's vote share in its *first* election to the U.S. Congress.
- Benchmark regressions: local linear regressions on a 3% bandwidth.
- We report our baseline regression estimates using 3 dependent variables:
  - Wealth in levels (rents as bribes or side-payments).
  - Wealth in logs (rents as returns on initial wealth).

Balance Tests

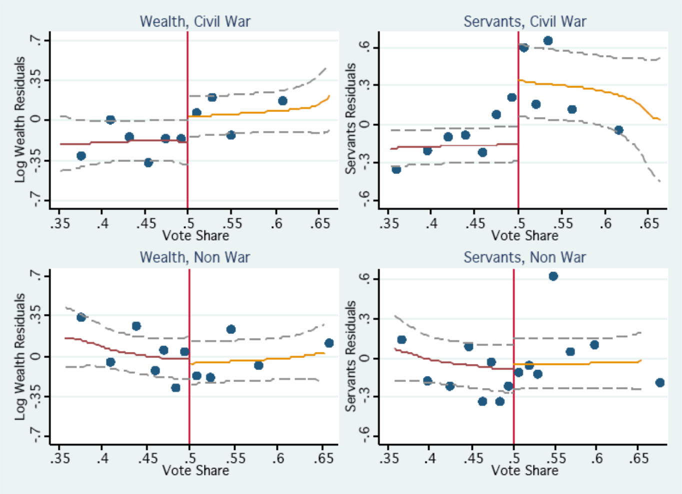
Table 10: Balance on Covariates in RDD Samples (3% margin)				
1850-1860 Period	Winner Mean	Loser Mean	Difference	p-Value
Log Initial Real Wealth	8.73	8.45	0.28	0.16
Log Initial Servants	0.72	0.62	0.10	0.52
Age	50.94	50.36	0.59	0.59
Lawyer Dummy	0.68	0.66	0.02	0.79
Manuf/Merch/Banker	0.18	0.20	-0.02	0.71
Farmer Dummy	0.22	0.18	0.04	0.46
1860-1870 Period				
Log Initial Total Wealth	9.72	9.67	0.05	0.79
Log Initial Servants	1.16	1.14	0.02	0.90
Age	41.21	41.49	-0.28	0.79
Lawyer Dummy	0.67	0.65	0.02	0.74
Manuf/Merch/Banker	0.19	0.25	-0.06	0.22
Farmer Dummy	0.17	0.14	0.03	0.49
Civil War Years				
Log Initial Total Wealth	10.10	9.79	0.05	0.23
Log Initial Servants	1.39	1.38	0.02	0.98
Age	42.93	42.64	-0.28	0.86
Lawyer Dummy	0.54	0.66	0.02	0.20
Manuf/Merch/Banker	0.28	0.21	-0.06	0.40
Farmer Dummy	0.26	0.20	0.03	0.39
Non-War Years				
Log Initial Total Wealth	9.28	9.58	0.05	0.28
Log Initial Servants	0.92	0.94	0.02	0.92
Age	38.78	40.43	-0.28	0.28
Lawyer Dummy	0.70	0.66	0.02	0.57
Manuf/Merch/Banker	0.17	0.26	-0.06	0.20
Farmer Dummy	0.13	0.10	0.03	0.53
1870-1880 Period				
Log Initial Servants	1.15	1.11	0.05	0.78
Age	34.85	37.08	-2.22	0.05
Lawyer Dummy	0.65	0.49	0.15	0.03
Manuf/Merch/Banker	0.25	0.28	-0.03	0.57
Farmer Dummy	0.15	0.28	-0.14	0.01

Summary of Results

- No evidence of large returns to congressional seats for the 1850's or the second half of the 1860's.
- However, congressmen who served during the **Civil War** accumulated on average, about 40% more wealth between 1860 and 1870 than candidates who lost the election and did not serve.
- Wealth accumulation was particularly significant by:
    - Congressmen who represented **states** that played an important role providing **supplies during the war**.
    - Congressmen who served during the Civil War in **committees** that were **responsible for most military appropriations**.

Findings

RDD Plots of Residuals from Regressions of Ending Log Wealth on Initial Log Wealth



Interpretation

- Sudden spike in government spending during the war and the decrease in oversight from government agencies might have made it easier for incumbent congressmen (and other politicians) to collect rents.
- More broadly, our results suggest that corruption and rent extraction may be more likely to occur in episodes of crisis (natural disasters, wars or other types of political and economic turmoil).
- During these periods government expenditure often increases substantially, increasing the amount of resources on which politicians might prey, and at the same time oversight by the media and other state institutions may be less effective than in normal times.

Federal Expenditure

Federal Government Spending Before, During and After the Civil War