

# Regressão descontínua – RD ou Desenho de regressão descontínua - RDD

Cunningham

- Requisito para RD estimar um efeito causal: hipótese de continuidade
- Os resultados potenciais mudam suavemente como uma função da *running variable* através do *cutoff*
- A única coisa que faz o resultado mudar subitamente no *cutoff* é o tratamento

Isso pode ser violado na prática:

1. A regra de atribuição é conhecida antecipadamente
2. Os agentes estão interessados em ajustar
3. Os agentes têm tempo para ajustar
4. O *cutoff* é endógeno a fatores que fazem os resultados potenciais mudarem
5. Há amontoamento não aleatório através da *running variable*

# Teste de *McCrary*

- Se assumimos um tratamento  $D$  e uma regra de atribuição  $X \geq c_0$ , então esperamos que os indivíduos irão se classificar em  $D$  escolhendo  $X$  tal que  $X \geq c_0$ , desde que eles sejam capazes de fazer isso
- Viés de seleção se classificação é uma função dos resultados potenciais

- Teste que verifica se há amontoamento (*bunching*) das unidades no *cutoff*. Necessário um teste de densidade
- McCrary(2008) sugere um teste formal onde sob a hipótese nula, a densidade deveria ser contínua no ponto de *cutoff*

- Particionar a variável de atribuição em *bins* e calcular as frequências (ou seja, o número de observações em cada *bin*)
- Usar as contagens de frequência como variável dependente numa regressão linear local

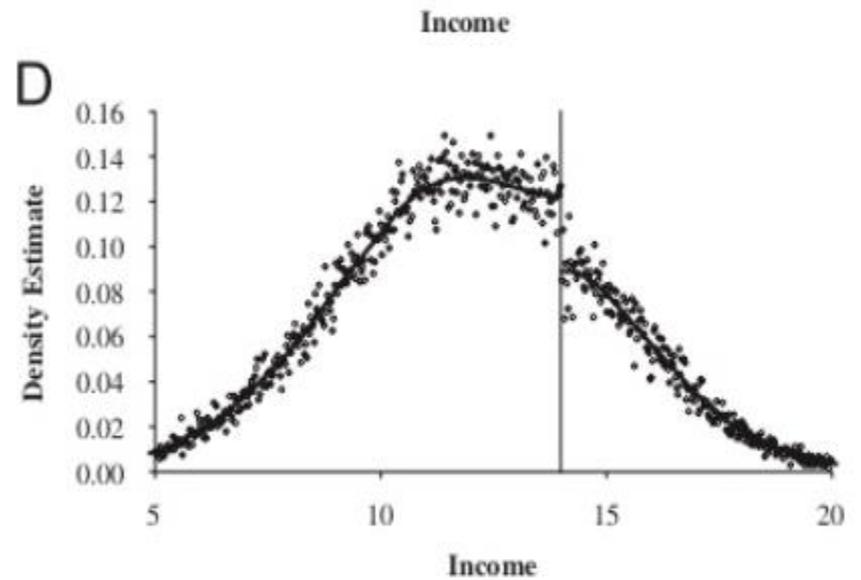
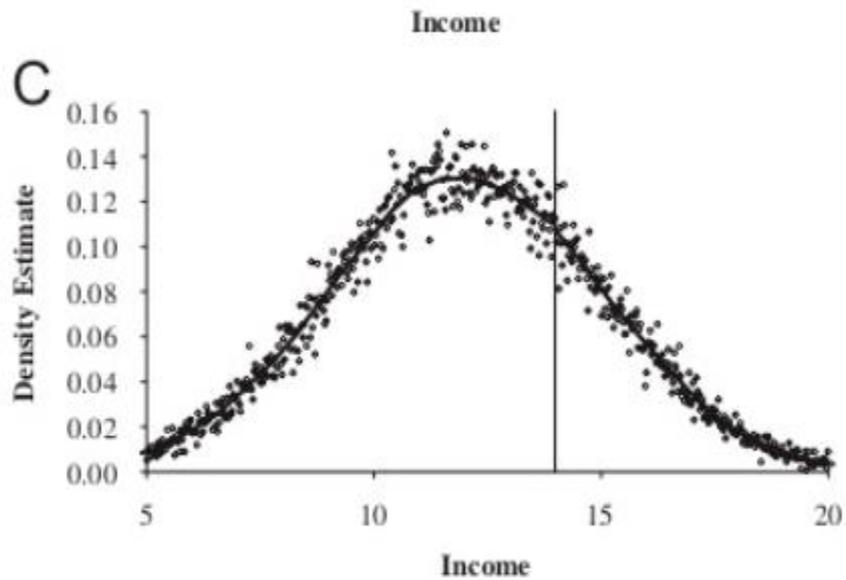


Figure 6.16: A picture with and without a discontinuity in the density from McCrary (2008).

## Balanceamento de covariadas e outros placebos

- Para o RDD ser válido, não deve haver uma mudança descontínua nos valores médios de covariadas em torno do *cutoff*
- Como essas são características pré-tratamento, elas deveriam ser invariantes a mudanças na atribuição ao tratamento
- Teste placebo: estamos procurando por ausência de efeito onde não deveria mesmo haver efeito

Lee, D.S., Moretti, E., Butler, M.J. (2004) Do voters affect or elect policies? Evidence from the U.S. House. *Quarterly Journal of Economics*, 119(3):807-859

## Os políticos ou os eleitores definem as políticas públicas?

- Teoria da convergência: políticos escolhem políticas moderadas para se adequar à heterogeneidade dos eleitores
- Teoria da divergência: políticos escolhem suas políticas preferidas, após tomarem posse; eleitores são incapazes de forçar um compromisso do político.

# Dados

- *Americans for Democratic Action (ADA)*
- Período: 1945-1990
- Unidade de análise: sessão distrital do congresso
- Número de observações: 13.588
- Variável dependente: **Score** (indicador de voto em políticas liberais); índice varia de 0 a 100. Maiores escores correspondem a um histórico de voto mais “liberal”
- Variável de tratamento: **Democrat** (variável binária indicando a vitória de democrata)
- *Running variable*: **Demvoteshare** (parcela de votos em democratas)

# Hipóteses

- Se a teoria da convergência for válida, então republicanos e democratas que venceram por uma diferença pequena deveriam votar de forma parecida. Se a teoria da divergência é verdadeira, eles deveriam votar de formas diferentes
- Se as preferências dos eleitores são as mesmas, mas as políticas divergem no ponto de corte, isso sugere que os políticos e não os eleitores estão conduzindo a formulação de políticas.

- Choque exógeno resulta da descontinuidade na *running variable*
- Se a parcela de votos é um pouco maior que 0,5, o candidato democrata ganha. O argumento é que em torno do *cutoff*, a vitória democrata foi determinada aleatoriamente, daí a atribuição aleatória de  $D_t$
- Somente a política do candidato vencedor é observada

## Especificações dos modelos paramétricos

- $\text{Score}_t = \beta_0 + \beta_1 \text{Democrat}_{t-1} + u_t$   
(especificação preferida)
- $\text{Score}_t = \beta_0 + \beta_1 \text{Democrat}_t + u_t$
- $\text{Democrat}_t = \beta_0 + \beta_1 \text{Democrat}_{t-1} + u_t$

## Tabela 46: Resultados baseados nos escores ADA – amostra de eleições apertadas

Dependent variable	$ADA_{t+1}$	$ADA_t$	$DEM_{t+1}$
Estimated gap	21.2	47.6	0.48
	(1.9)	(1.3)	(0.02)

Standard errors in parentheses. The unit of observation is a district-congressional session. The sample includes only observations where the Democrat vote share at time  $t$  is strictly between 48 percent and 52 percent. The estimated gap is the difference in the average of the relevant variable for observations for which the Democrat vote share at time  $t$  is strictly between 50 percent and 52 percent and observations for which the Democrat vote share at time  $t$  is strictly between 48 percent and 50 percent. Time  $t$  and  $t + 1$  refer to congressional sessions.  $ADA_t$  is the adjusted ADA voting score. Higher ADA scores correspond to more liberal roll-call voting records. Sample size is 915

- Uma vitória Democrata aumenta o voto liberal no próximo período em 21 pontos
- Uma vitória Democrata aumenta o voto liberal no período corrente em 48 pontos
- Uma vitória Democrata aumenta a probabilidade de reeleição em 48%
- Evidência para divergência e vantagem do incumbente

## Tabela 47: Resultados baseados nos escores ADA – amostra completa

Dependent variable	$ADA_{t+1}$	$ADA_t$	$DEM_{t+1}$
Estimated gap	31.50 <sup>***</sup>	40.76 <sup>***</sup>	0.82 <sup>***</sup>
	(0.48)	(0.42)	(0.01)
N	13,588	13,588	13,588

Cluster robust standard errors in parenthesis. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Tabela 48: Resultados baseados nos escores ADA – amostra completa

Tabela 46 e Tabela 47: regressões não usam controles  
Também não usam a recentralização da *running variable*  
Agora fazer os dois: subtrair 0,5 da *running variable*

Dependent variable	$ADA_{t+1}$	$ADA_t$	$DEM_{t+1}$
Estimated gap	33.45***	58.50***	0.55**
	(0.85)	(0.66)	(0.01)
N	13,577	13,577	13,577

Cluster robust standard errors in parenthesis. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Tabela 49: Resultados baseados em escores ADA – amostra completa com interações lineares

Interação da *running variable* com a variável de tratamento

Dependent variable	$ADA_{t+1}$	$ADA_t$	$DEM_{t+1}$
Estimated gap	30.51***	55.43***	0.53***
	(0.82)	(0.64)	(0.01)
N	13,577	13,577	13,577

Cluster robust standard errors in parenthesis. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Tabela 50: Resultados baseados nos escores ADA – amostra completa com interações lineares e quadráticas

Interação da *running variable* com a variável de tratamento e termo quadrático

Dependent variable	$ADA_{t+1}$	$ADA_t$	$DEM_{t+1}$
Estimated gap	13.03***	44.40 ***	0.32**
	(1.27)	(0.91)	(1.74)
N	13,577	13,577	13,577

Cluster robust standard errors in parenthesis. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## Tabela 51: Resultados baseados nos escores ADA – eleições apertadas com interações lineares e quadráticas

Limitar novamente a análise a uma janela menor

+/- 0,05 ao invés de +/- 0,02 como anteriormente

Agora estamos usando mais observações longe do *cutoff* e por isso temos 2.441 observações e não 915

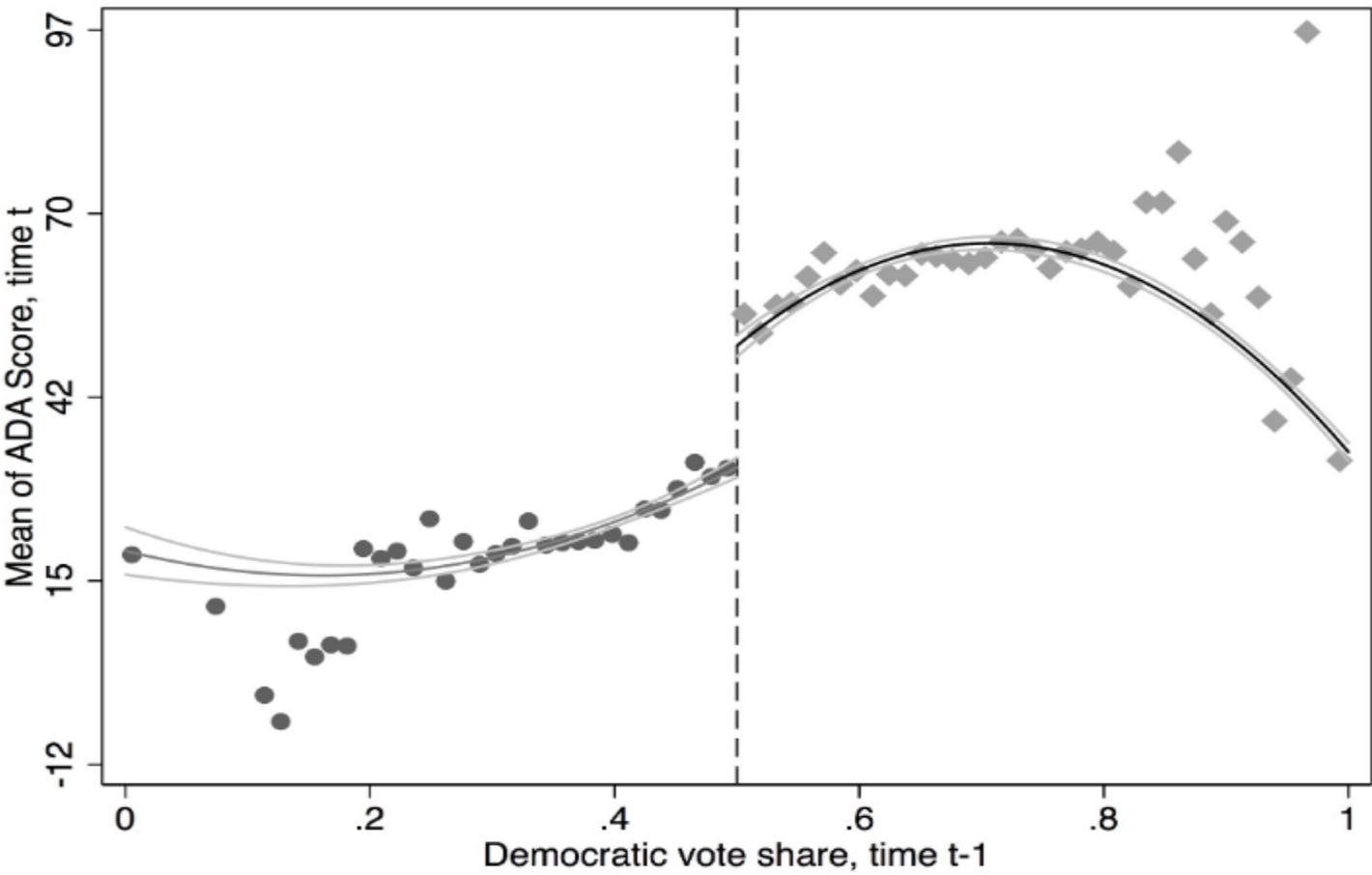
Dependent variable	$ADA_{t+1}$	$ADA_t$	$DEM_{t+1}$
Estimated gap	3.97***	46.88***	0.12**
	(1.49)	(1.54)	(0.02)
N	2,441	2,441	2,441

Cluster robust standard errors in parenthesis. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

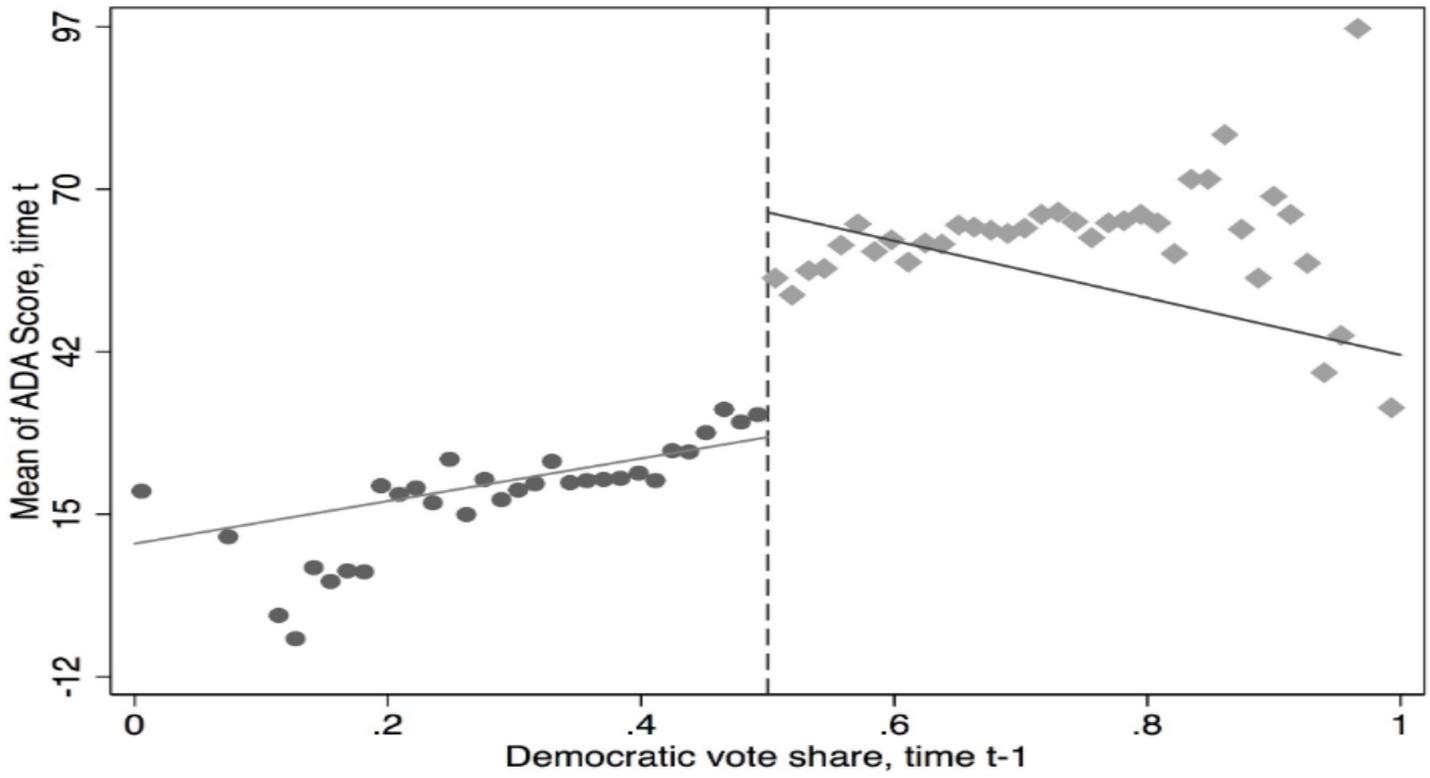
# Resumo

- 1) Ajuste do modelo sem controlar para a *running variable*
- 2) *Running variable* foi incluída de diferentes formas. Por exemplo, interação do *vote share* com a *dummy* para democrata, inclusão do termo quadrático
- 3) Inclusão da *running variable* em qualquer forma tendeu a reduzir o efeito da vitória de Democratas sobre os padrões futuros de voto dos Democratas
- 4) Efeito imediato permanece bastante grande

# Figura 38: ajuste cuadrático



# Figura 39: ajuste linear



# Hipótese de continuidade

- Envolve funções de expectativa condicional dos resultados potenciais contínuas através do *cutoff* e, portanto, não é testável
- O que é possível fazer: testar se existem mudanças nas funções de expectativa condicional para outras covariadas exógenas que não poderiam ou não deveriam mudar como resultado do *cutoff*
- Placebos: testes indiretos da hipótese de continuidade

# Seleção da janela

- Vimos a importância da seleção da *bandwidth* para a estimação do efeito causal, assim como a importância da seleção da extensão do polinômio
- Sempre existe um trade-off entre viés e variância ao escolher a *bandwidth*
- Seleção ótima de bandwidth Calonico et al. (2014)

[lmb 9.R](#)

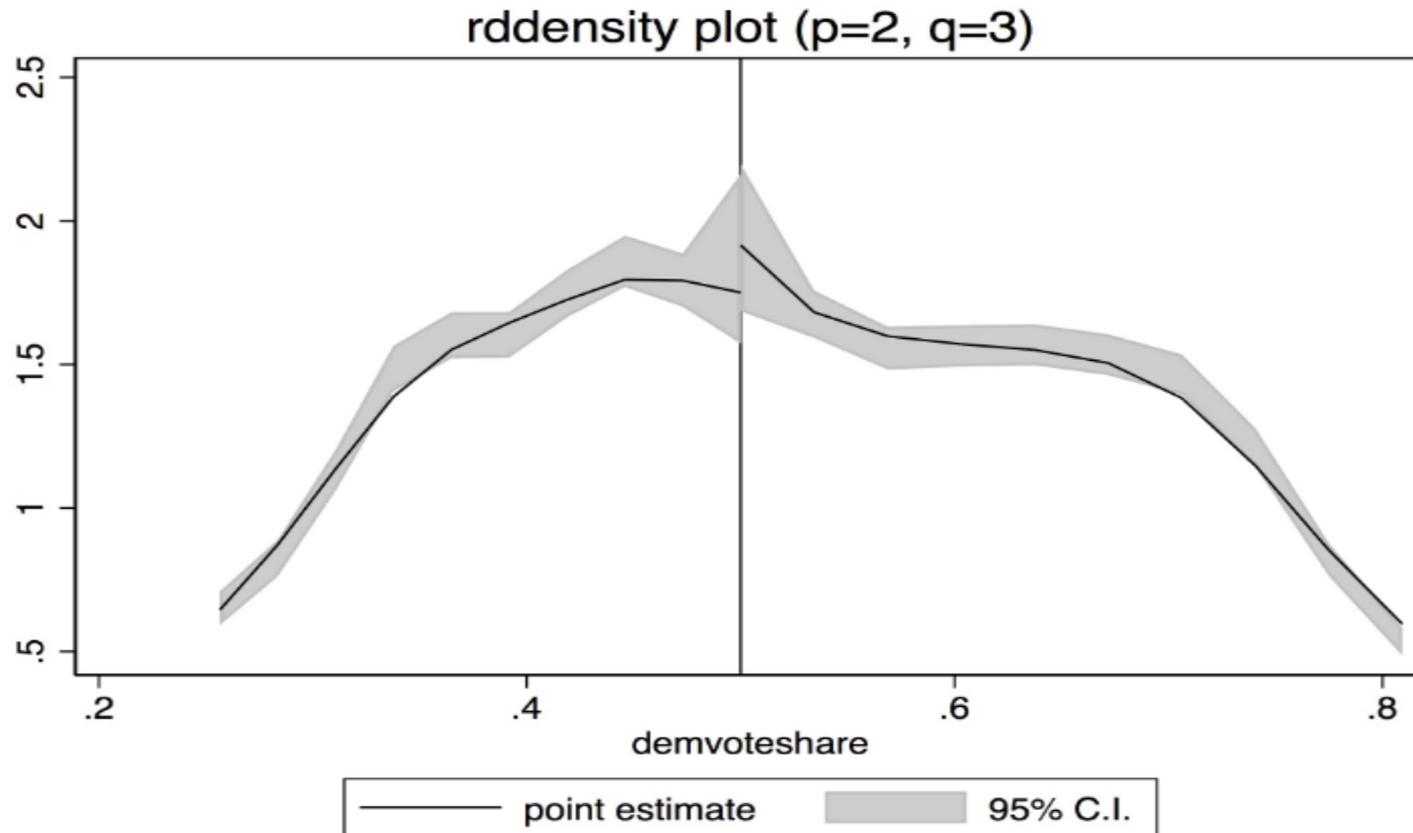
▼ Code

```
library(tidyverse)
library(rdrobust)

rdr <- rdrobust(y = lmb_data$score,
               x = lmb_data$demvoteshare, c = 0.5)

summary(rdr)
```

# Teste de McCrary



- Todos os Scripts estão no Moodle, assim como uma guia para reproduzir o que foi dado nesta aula e mais