"Notes on Experiments and quasi-experiments in Economics"
São Paulo School of Advanced Science on Learning from Data
07/Aug/2019

Sergio Firpo
Insper

Introduction

- How to determine if there is a Causal Effect of a Program on Some Variable of Interest.

- Example 1: Workers Training / Qualification Program:

    - Workers undergoing the program are compared with non-program workers. In general, can it be argued that the average difference in labor income between the groups is unequivocally due to the training effect?

- Example 2: Conditional cash transfer programs:

- Children whose families belong to the program are compared to children who do not belong to the program. In general, it can be said that the average difference in education (measured in some way, such as matricule, lag, number of failures) between the decreasing groups is due, unequivocally, to the effect of the program?

- Answer for both examples: **NO**

- What are the conditions under which it can be determined if there is a causal effect?
- Evaluation Design:
  - Groups chosen in a random way;
  - Groups chosen based on observables;
  - Groups are result of self-selection;
- Assumptions about selection into treatment:
  - Researcher has access to observable determinants of selection;
  - Researcher does not have access to important determinants of selection.

Plan of this talk

- Introduction: Identifying Program Effect, Causality, Potential Results, Heterogeneity, and Selection
- Random Experiments: Advantages, Disadvantages, and Examples
- Selection based on Observables
  - Basic Concepts
  - Matching
  - Propensity-Score Methods
- Selection based on Non-Observables
  - Instrumental Variables
  - Regression-Discontinuity Designs
  - Synthetic Control Methods

Identification

- Identifying β in the equation below

$$Y_i = \alpha + \beta \cdot T_i + U_i$$

- where Y is outcome variable and T treatment assignment dummy.

Causality

- Causality for us is related to possibility of manipulation or of taking an action.
- Examples:
  - "I took a pill and my headache disappeared."
  - "She got a nice job because she graduated from an elite school."
  - "He got fired because is old."

Potential outcomes and treatment effect heterogeneity

- Potential Outcomes:
  $Y_i(0)$, outcome had i not received the treatment;
  $Y_i(1)$, outcome had i received the treatment;

- We are interested in :

$$\beta_i = Y_i(1) - Y_i(0)$$

Central problem of causal inference (Holland, 1986):

- We cannot observe both $Y_i(1)$ and $Y_i(0)$ simultaneously for same unit i.
- Thus, we cannot observe $\beta_i = Y_i(1) - Y_i(0)$

- We observe Y_i:

$$\begin{aligned} Y_i &= T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0) \\ &= Y_i(0) + (Y_i(1) - Y_i(0)) \cdot T_i \end{aligned}$$

- If we could observe both potential outcomess:
- We would then have individual treatment effects:

$$\beta_i = Y_i(1) - Y_i(0)$$

- And we would estimate the *average treatment effect*, ATE by

$$\frac{1}{N}\sum_{i=1}^{N}\left(Y_i(1) - Y_i(0)\right) \xrightarrow{P} E\left[Y(1) - Y(0)\right] = ATE$$

- We observe nevertheless $\{Y_i, T_i, X_i\}_{i=1}^{N}$ where X_i is a vector of covariates.

- From the definition of Y, we have

$$\begin{aligned} Y_i &= \alpha_i + \beta_i \cdot T_i \\ &= \alpha + \beta \cdot T_i + U_i \end{aligned}$$

where

$$U_i = \alpha_i - \alpha + (\beta_i - \beta) \cdot T_i$$

$$\begin{aligned} \alpha &= E[\alpha_i] \\ \beta &= E[\beta_i] = ATE \end{aligned}$$

- and if

$$\begin{aligned} & T_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) \\ \Rightarrow \quad & E[U_i | T_i] = Cov(T_i, U_i) = 0 \end{aligned}$$

- Sometimes we are interested in the ATT, or the average treatment effect on the treated:

$$ATT = E\left[Y\left(1\right) - Y\left(0\right)|T = 1\right]$$

- For the case of random experiments

$$T_i \perp\!\!\!\perp \left(Y_i\left(1\right), Y_i\left(0\right)\right)$$

then $ATT = ATE$

Selection based on observables:

- Given X, potential outcomes are independent of the treatment:

$$\Pr\left[T=1|Y\left(1\right),Y\left(0\right),X\right]=\Pr\left[T=1|X\right]\equiv p\left(X\right)$$

where $p\left(X\right)$, or the probability of being treated given X, is the so-called **propensity-score**.

- Conditional independence assumption (given X) between T and the potential outcomes $Y(1)$ and $Y(0)$ a.k.a *treatment ignorability* or *unconfoundedness*.

T is ignorable if

$$\Pr[T = 1|Y(1), Y(0), X] = \Pr[T = 1|X]$$

or

$$(Y(1), Y(0)) \perp\!\!\!\perp T|X$$

- Given ignorability, there are several ways to identify ATE from data.

- Imputation:

$$\begin{aligned} ATE &= E\left[Y\left(1\right)\right] - E\left[Y\left(0\right)\right] \\ &= E\left[E\left[Y\left(1\right)|X\right]\right] - E\left[E\left[Y\left(0\right)|X\right]\right] \\ &= E\left[E\left[Y\left(1\right)|X, T=1\right]\right] - E\left[E\left[Y\left(0\right)|X, T=0\right]\right] \\ &= E\left[E\left[Y|X, T=1\right]\right] - E\left[E\left[Y|X, T=0\right]\right] \end{aligned}$$

- If we are interested in ATT, then

$$\begin{aligned} ATT &= E\left[Y\left(1\right)|T=1\right]-E\left[Y\left(0\right)|T=1\right] \\ &= E\left[Y|T=1\right]-E\left[E\left[Y\left(0\right)|X,T=1\right]|T=1\right] \\ &= E\left[Y|T=1\right]-E\left[E\left[Y\left(0\right)|X,T=0\right]|T=1\right] \\ &= E\left[Y|T=1\right]-E\left[E\left[Y|X,T=0\right]|T=1\right] \end{aligned}$$

- Thus, we can estimate ATE and ATT by:

$$\widehat{ATE}_{imp} = \frac{1}{N}\sum_{i=1}^{N}\widehat{E}\left[Y|X_i, T=1\right] - \frac{1}{N}\sum_{i=1}^{N}\widehat{E}\left[Y|X_i, T=0\right]$$

$$\widehat{ATT}_{imp} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{\widehat{p}}\cdot\left(Y_i - \widehat{E}\left[Y|X_i, T=0\right]\right)$$

Matching

- For individual i, if $Y_i(1)$ observed $(T_i = 1)$, then $Y_i(0)$ is *missing*.
- Find j, such that $T_j = 0$ and:

$$\|X_j - X_i\| \leq \|X_l - X_i\|, \qquad \forall\, l,\ T_l = 0$$

- Then $\widehat{Y}_i(0) = Y_j(0) = Y_j$.

- ATT is estimated by

$$\widehat{ATT}_{match,1} = \frac{1}{N}\sum_{i=1}^{N} \frac{T_i}{\widehat{p}} \cdot \left(Y_i - \widehat{Y}_i(0)\right)$$

- Implementation questions:

1. How to deal with ties?
2. With or without replacement?
3. What distance $\|\cdot\|$ should be used?
4. $\|X_i - X_j\|$ is different from zero, in general. That generates bias. How to deal with that? Bias increases with dimension of X.
5. Statistical inference.
6. Number of matches: bias-variance tradeoff
7. Bootstrap fails

- Finally, note that one interpretation for matching is as a particular case of imputation method. For ATT:

$$\widehat{ATT}_{imp} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{\widehat{p}}\cdot\left(Y_i - \widehat{E}\left[Y|X_i, T=0\right]\right)$$

$$\widehat{ATT}_{match,1} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{\widehat{p}}\cdot\left(Y_i - \widehat{Y}_i\left(0\right)\right)$$

and difference between $\widehat{E}\left[Y|X_i, T=0\right]$ and $\widehat{Y}_i\left(0\right)$ may not even exist. If, for example, $E\left[Y|X, T=0\right]$ is estimated by nearest-neighbor with fixed number of neighbors, these methods are algebraically identical.

Inverse probability weighting

$$\begin{aligned} ATE &= E\left[E\left[Y|X,T=1\right]\right] - E\left[E\left[Y|X,T=0\right]\right] \\ &= E\left[\frac{T}{p\left(X\right)}\cdot Y\right] - E\left[\left(\frac{1-T}{1-p\left(X\right)}\right)\cdot Y\right] \end{aligned}$$

$$\begin{aligned} ATT &= \frac{E\left[T\cdot Y\right]}{p} - \frac{E\left[p\left(X\right)\cdot E\left[Y|X,T=0\right]\right]}{p} \\ &= E\left[\frac{T}{p}\cdot Y\right] - E\left[\left(\frac{p\left(X\right)}{p}\cdot\frac{1-p}{1-p\left(X\right)}\cdot\frac{1-T}{1-p}\right)\cdot Y\right] \end{aligned}$$

- Selection on Unobservables: Instrumental Variables Method

- How to estimate ATE when

$$\Pr\left[T=1|Y\left(1\right),Y\left(0\right),X\right]\neq\Pr\left[T=1|X\right]$$

- Let Z be a dummy instrumental variable.

- Define potential treatment $T_i(z)$, as the treatment that individual i would have had she got $Z = z$. Observed treatment is then:

$$T_i = T_i(1) \cdot Z_i + T_i(0) \cdot (1 - Z_i)$$

- And observed outcome still is:

$$Y_i = Y_i(1) \cdot T_i + Y_i(0) \cdot (1 - T_i)$$

- Identification assumptions:
  - Instrument validity:

$$(Y(1), Y(0), T(1), T(0)) \perp\!\!\!\perp Z$$

  - Monotonicity:

$$T(1) \geq T(0)$$

- Four types, but one is ruled out (*defiers*):

$T(1)$	$T(0)$	type
1	1	*always-taker* (a)
1	0	*complier* (c)
0	1	*defier* (d)
0	0	*never-taker* (n)

- One can obtain an average treatment effect that is 'local', as it is an ATE for "compliers", or simply, $LATE$:

$$\begin{aligned} LATE &= E\left[Y\left(1\right)-Y\left(0\right)|T\left(1\right)>T\left(0\right)\right] \\ &= \frac{E\left[Y|Z=1\right]-E\left[Y|Z=0\right]}{E\left[T|Z=1\right]-E\left[T|Z=0\right]} \end{aligned}$$

Regression Discontinuity Design

- Novamente, o ideal seria estimar $ATE = E\left[Y\left(1\right) - Y\left(0\right)\right]$. Lembre-se que:

$$\begin{aligned} Y_i &= Y_i\left(0\right) + \left(Y_i\left(1\right) - Y_i\left(0\right)\right) \cdot T_i \\ &= \alpha + \beta \cdot T_i + Y_i\left(0\right) - \alpha + \left(Y_i\left(1\right) - Y_i\left(0\right) - \beta\right) \cdot T_i \\ &= \alpha + \beta \cdot T_i + U_i \end{aligned}$$

onde

$$\begin{aligned} \beta &= ATE \\ U_i &= Y_i\left(0\right) - \alpha + \left(Y_i\left(1\right) - Y_i\left(0\right) - \beta\right) \cdot T_i \end{aligned}$$

- Say that there is some continuous variable Z, sucht that $E[T|Z]$ is well defined.

- There is a point z_0, in which the probability of being treated jumps:

$$\begin{aligned} T^+ &\equiv \lim_{\epsilon \downarrow 0} \Pr\left[T = 1 | Z = z_0 + \epsilon\right] \\ T^- &\equiv \lim_{\epsilon \downarrow 0} \Pr\left[T = 1 | Z = z_0 - \epsilon\right] \\ T^+ &\neq T^- \end{aligned}$$

- This is the source of exogenous variation that allows us to identify ATE at z_0.

- *Fuzzy design* versus *sharp design*

  - sharp design

$$\begin{aligned} Pr[T &= 1|Z < z_0] = 0 \\ Pr[T &= 1|Z > z_0] = 1 \end{aligned}$$

  - fuzzy design:

$$\begin{aligned} 0 &< Pr[T = 1|Z < z_0] < 1 \\ 0 &< Pr[T = 1|Z > z_0] < 1 \end{aligned}$$

- Assumption: $E[Y(1)|Z=z]$ and $E[Y(0)|Z=z]$ are continuous at $z=z_0$.

- Define

$$\begin{aligned} Y^+ &\equiv \lim_{\epsilon \downarrow 0} E\left[Y|Z=z_0+\epsilon\right] \\ Y^- &\equiv \lim_{\epsilon \downarrow 0} E\left[Y|Z=z_0-\epsilon\right] \end{aligned}$$

- Then ATE at z_0 is

$$\begin{aligned} ATE(z_0) &= E\left[Y\left(1\right) - Y\left(0\right)|Z = z_0\right] \\ &= \frac{Y^+ - Y^-}{T^+ - T^-} \end{aligned}$$

Figure 1: Simple Linear RD Setup

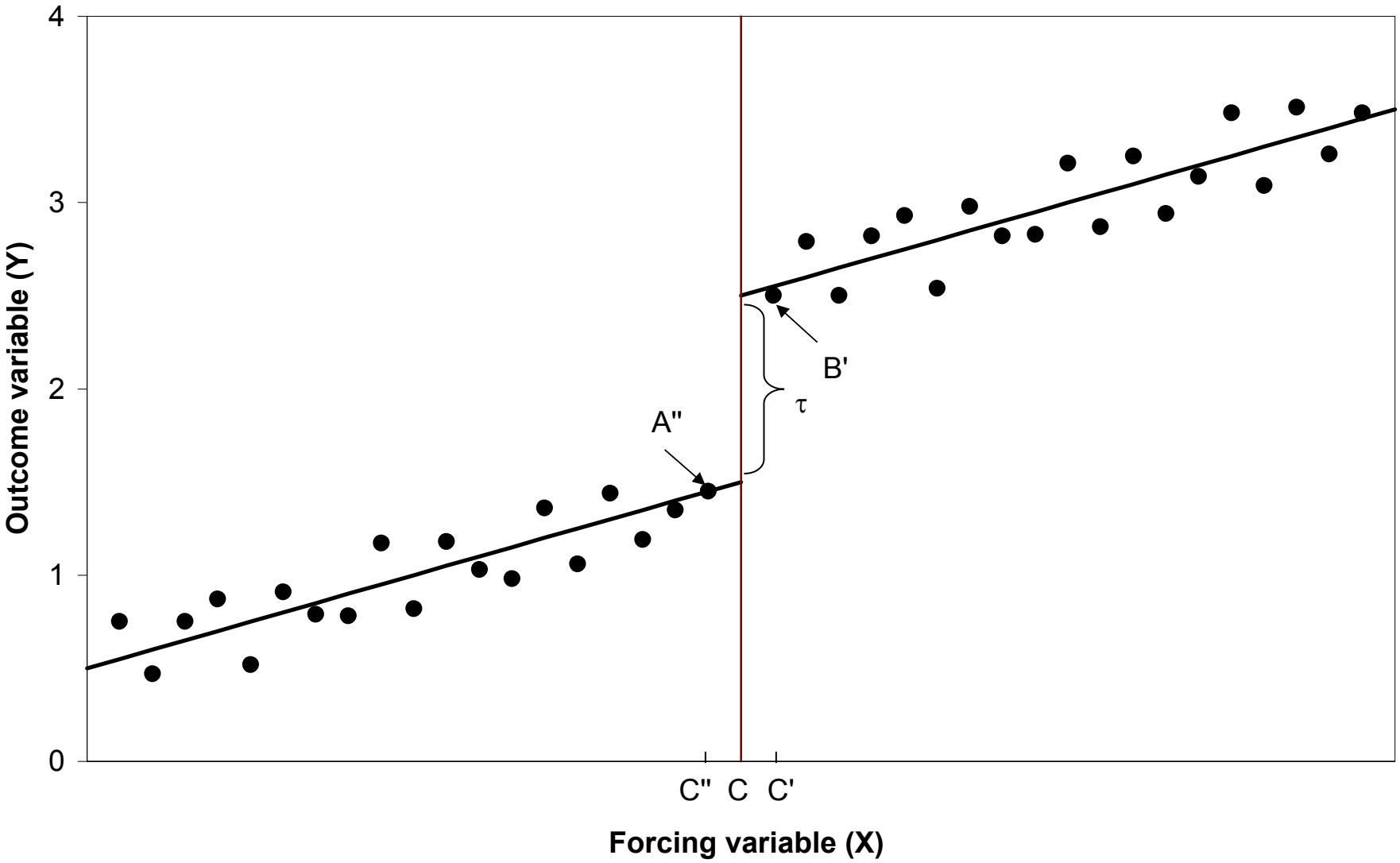


Figure 2: Nonlinear RD

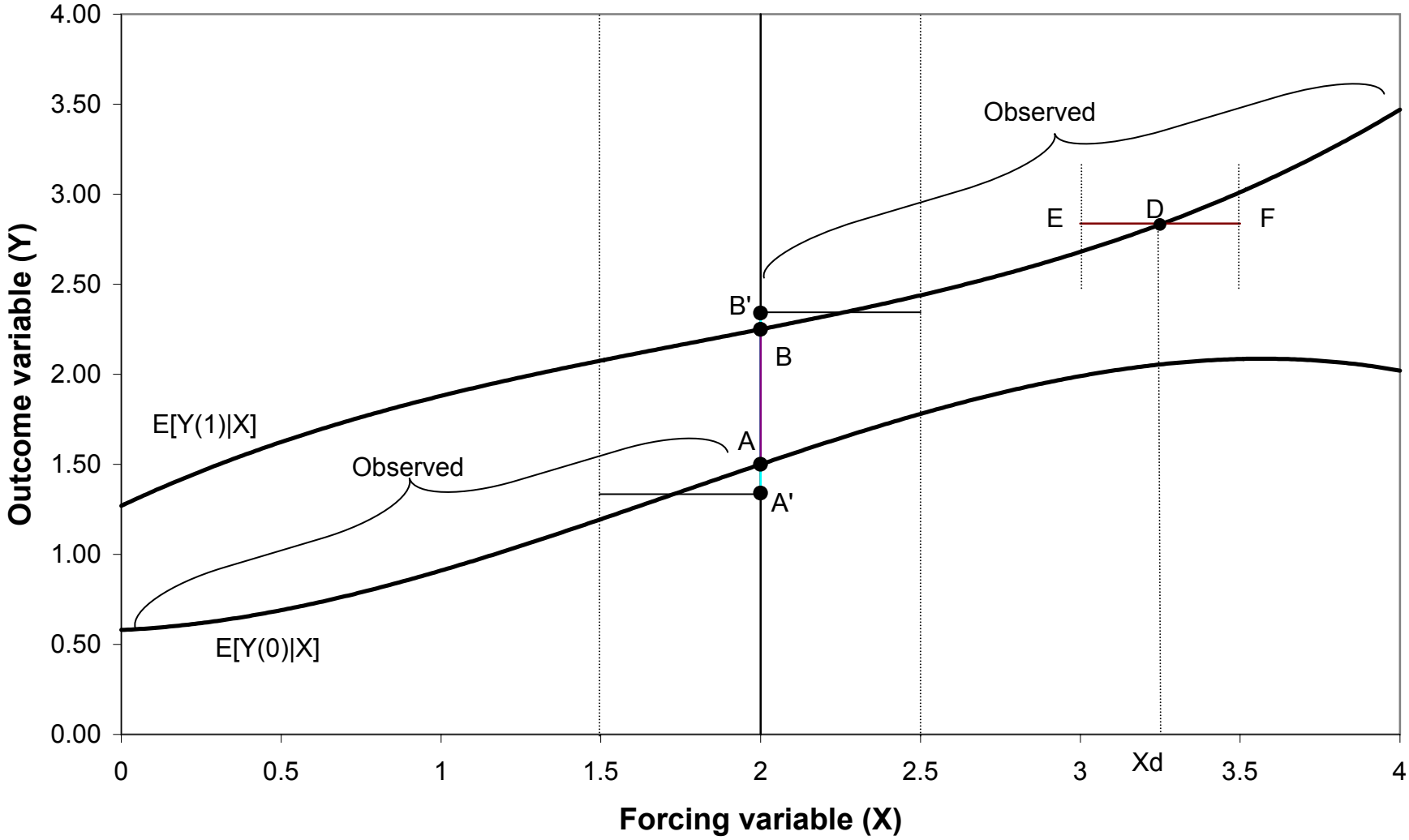


Figure 3: Randomized Experiment as a RD Design

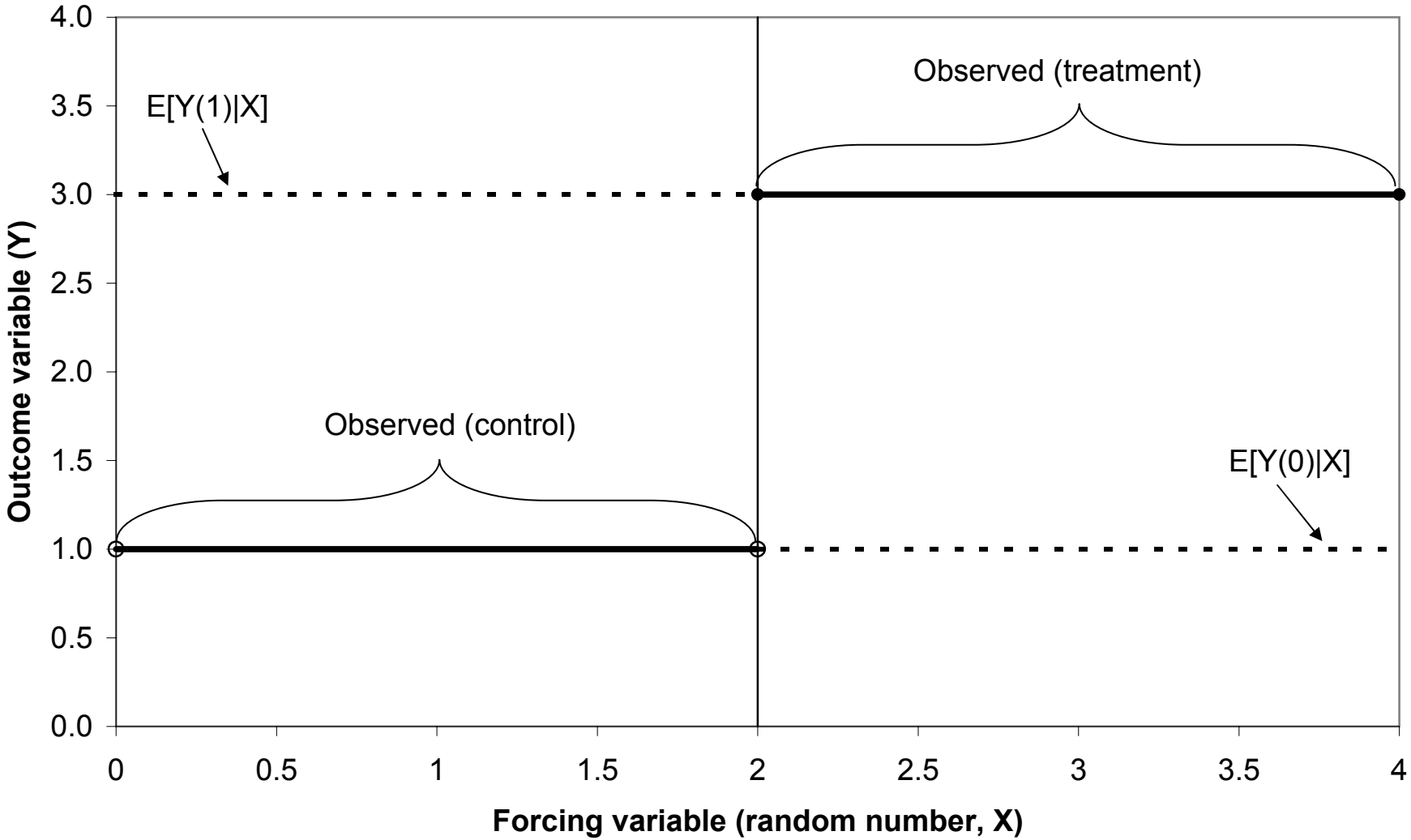


Figure 4: Density of Forcing Variable Conditional on W=w, U=u

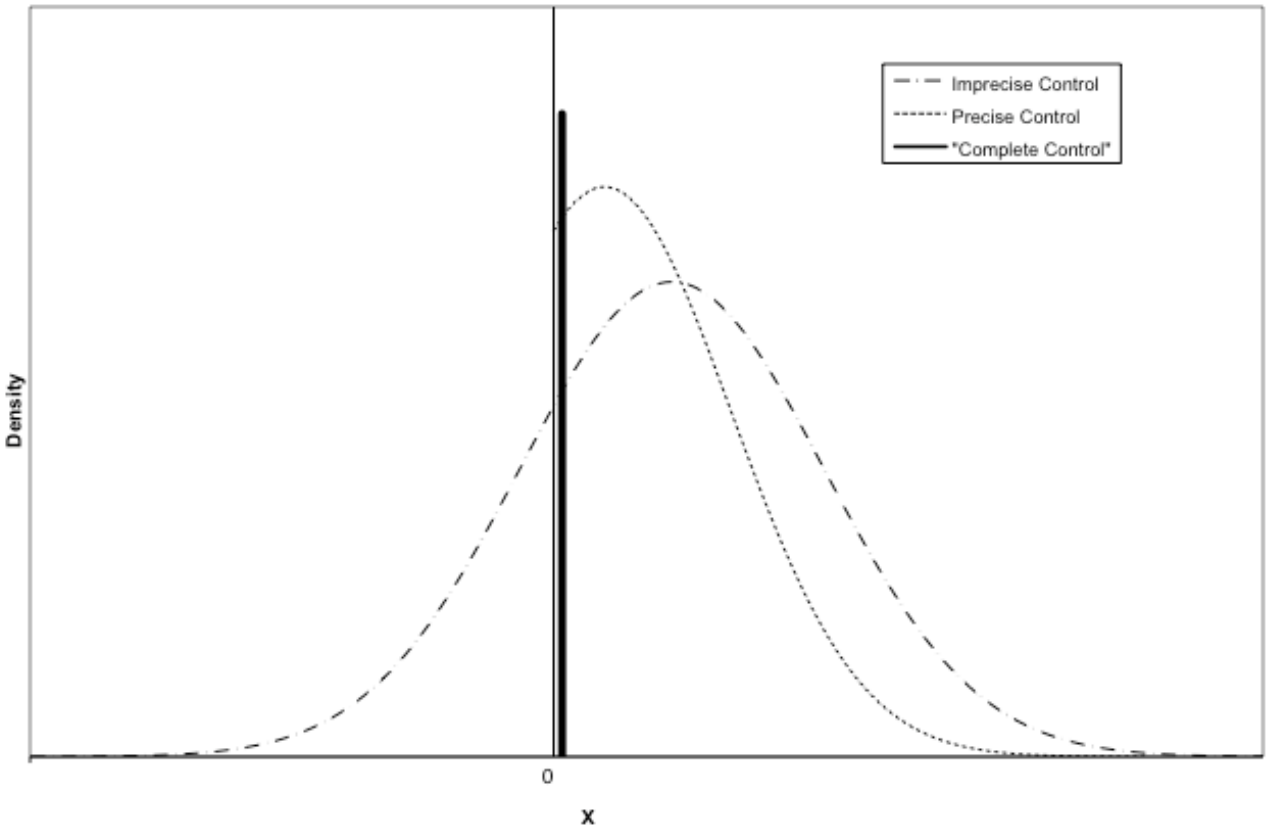
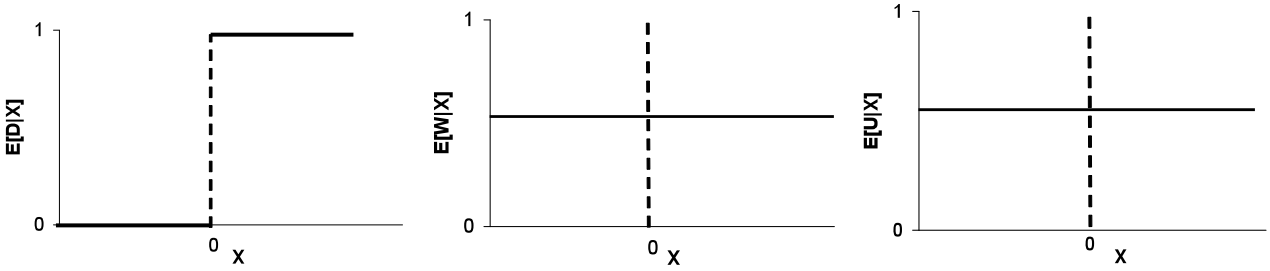
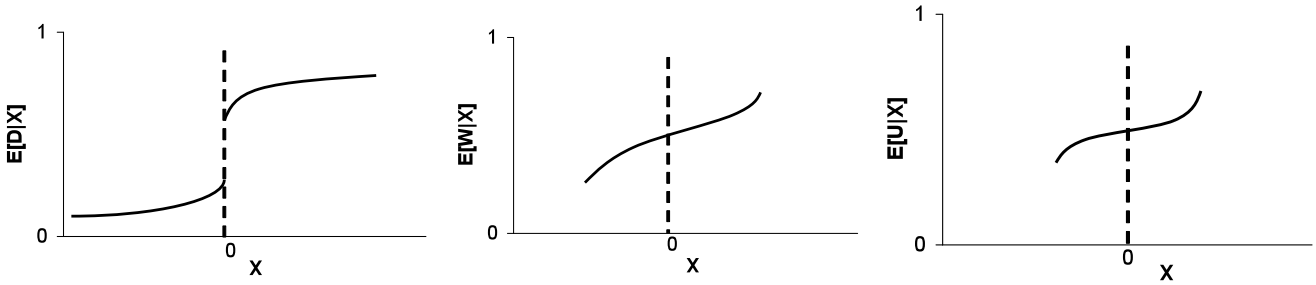


Figure 5. Treatment, Observables, and Unobservables in four research designs.

A. Randomized Experiment



B. Regression Discontinuity Design



C. Matching on Observables

D. Instrumental Variables

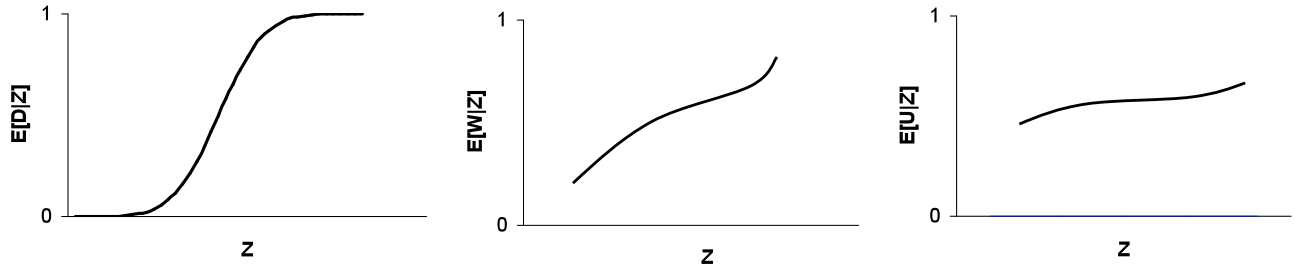


Figure 6a: Share of vote in next election, bandwidth of 0.02 (50 bins)

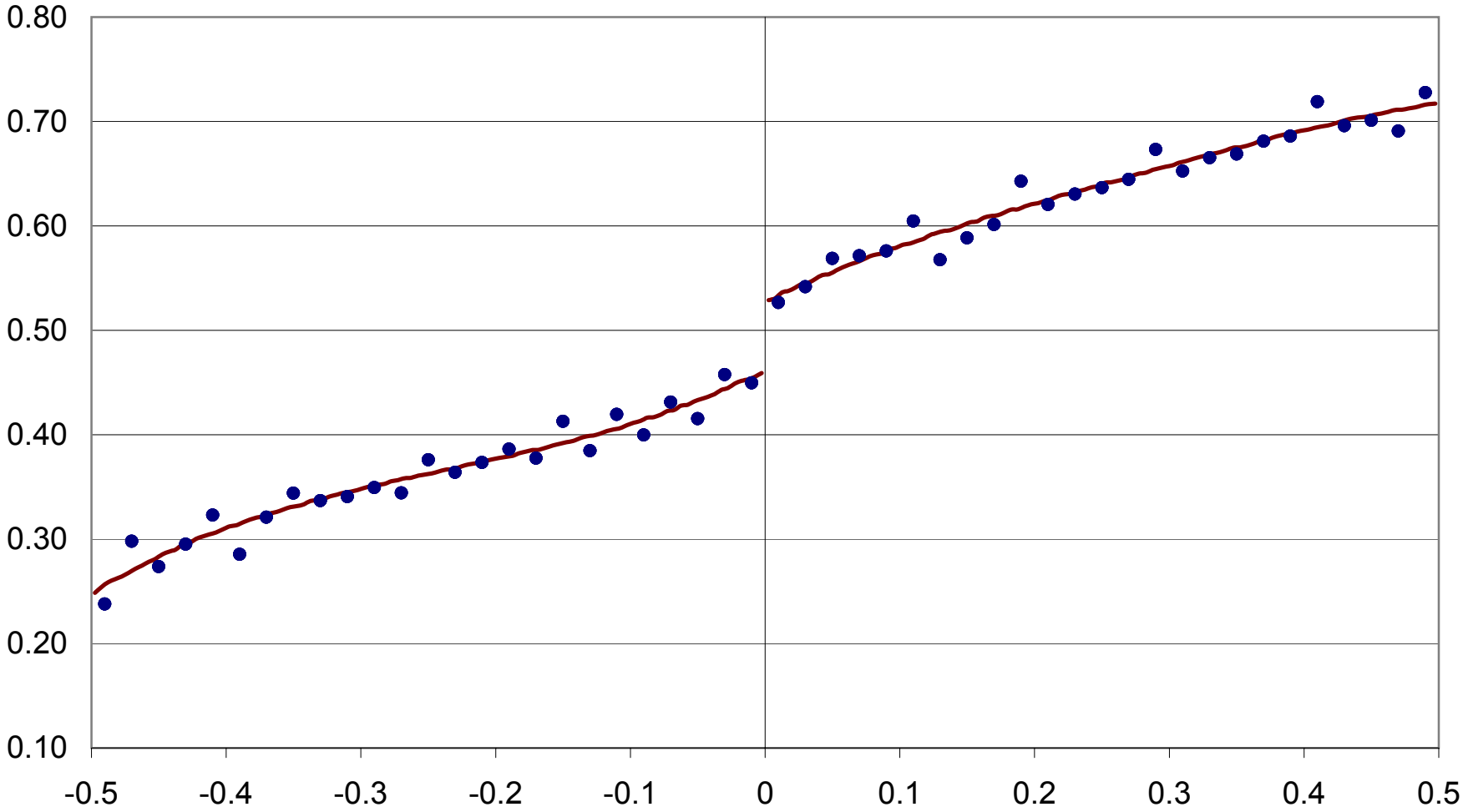


Figure 6b: Share of vote in next election, bandwidth of 0.01 (100 bins)

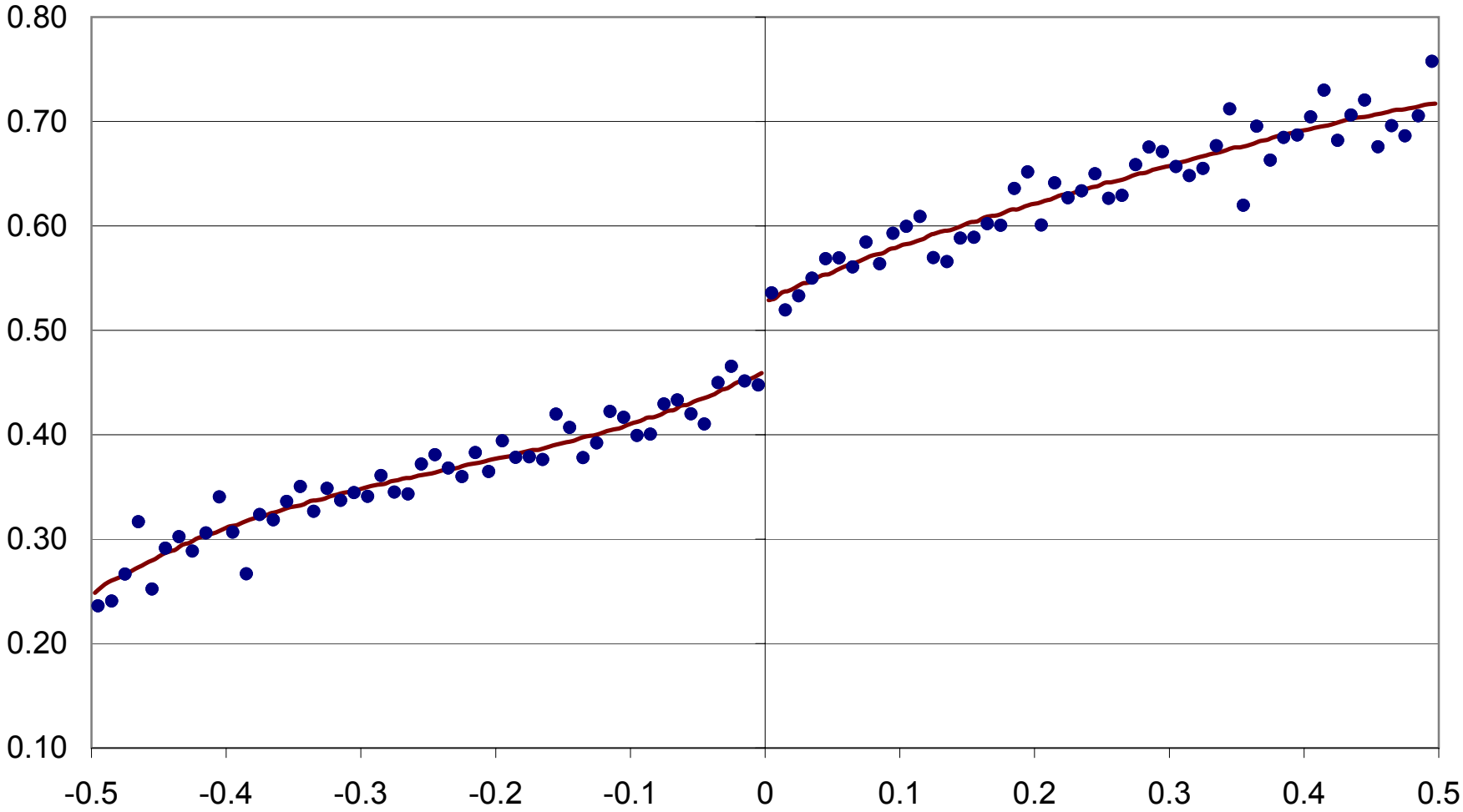


Figure 6c: Share of vote in next election, bandwidth of 0.005 (200 bins)

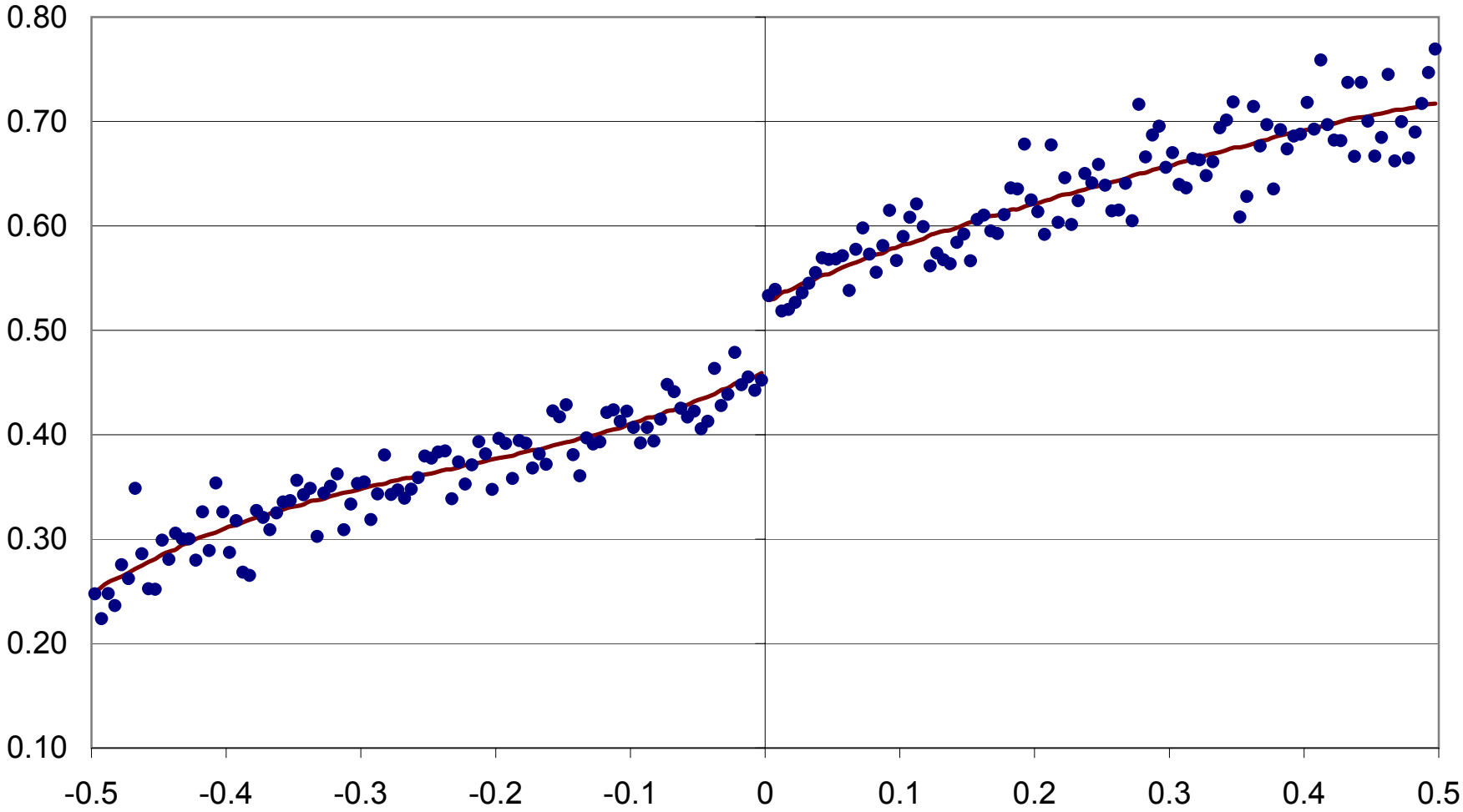


Figure 7a: Winning the next election, bandwidth of 0.02 (50 bins)

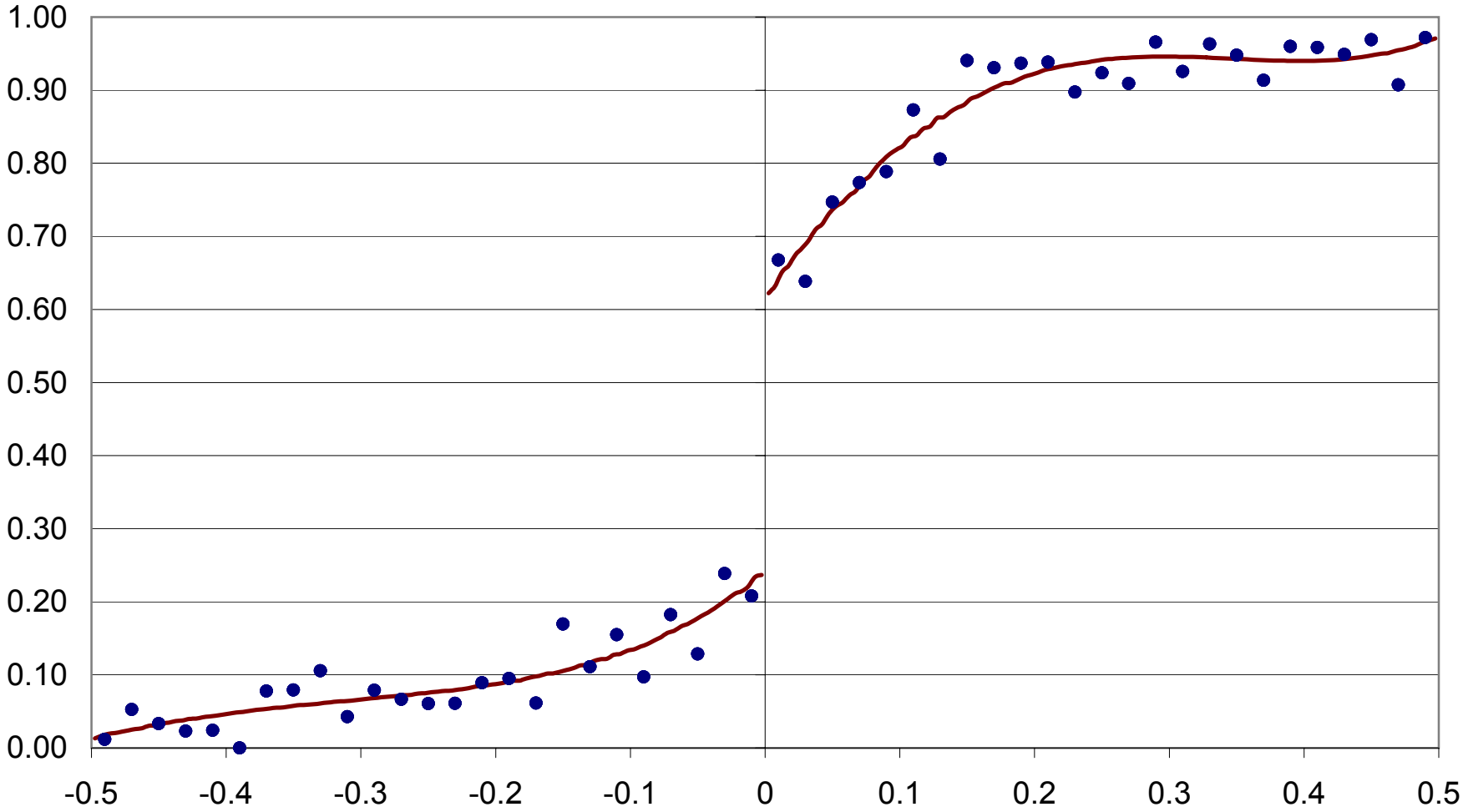


Figure 7b: Winning the next election, bandwidth of 0.01 (100 bins)

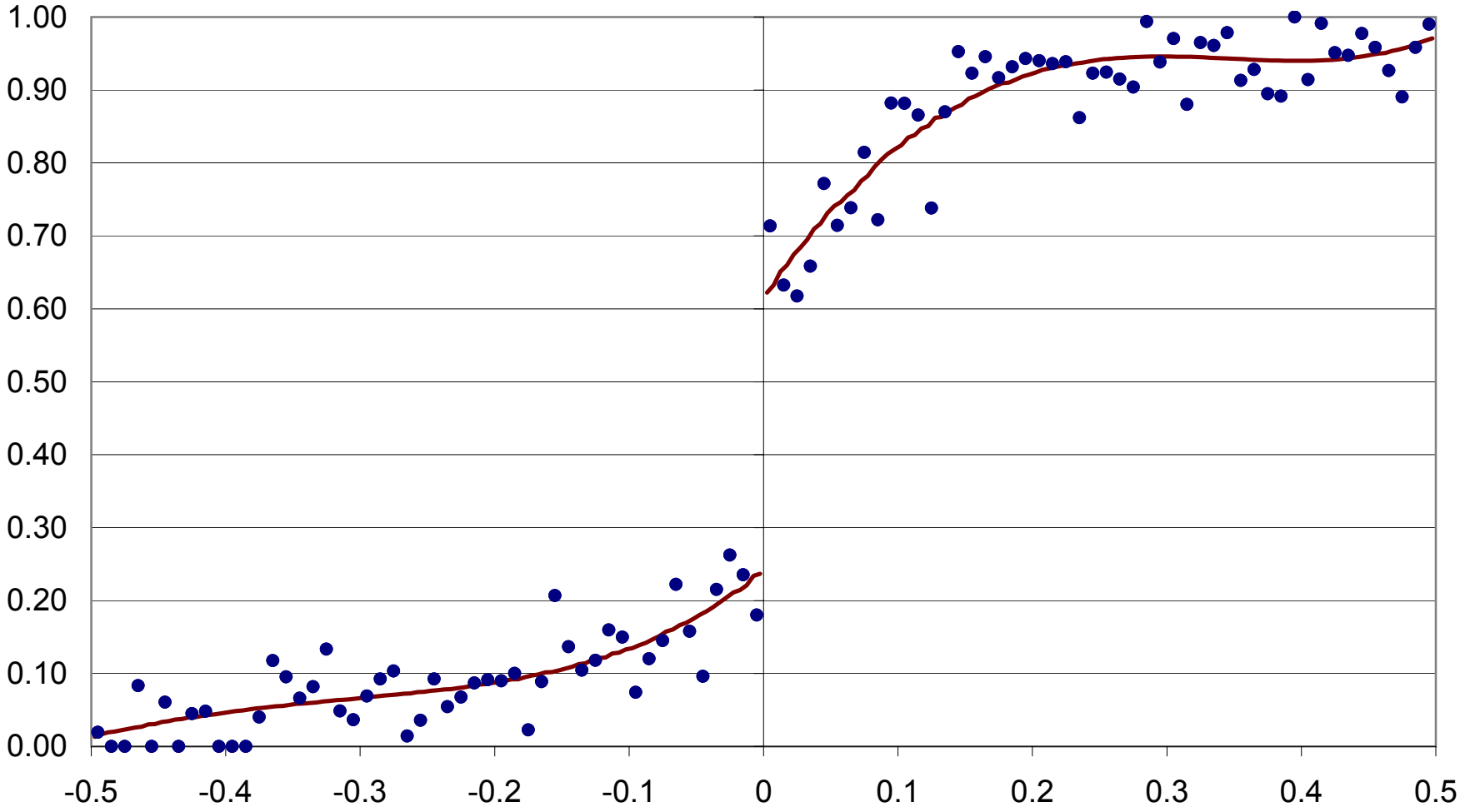


Figure 7c: Winning the next election, bandwidth of 0.005 (200 bins)

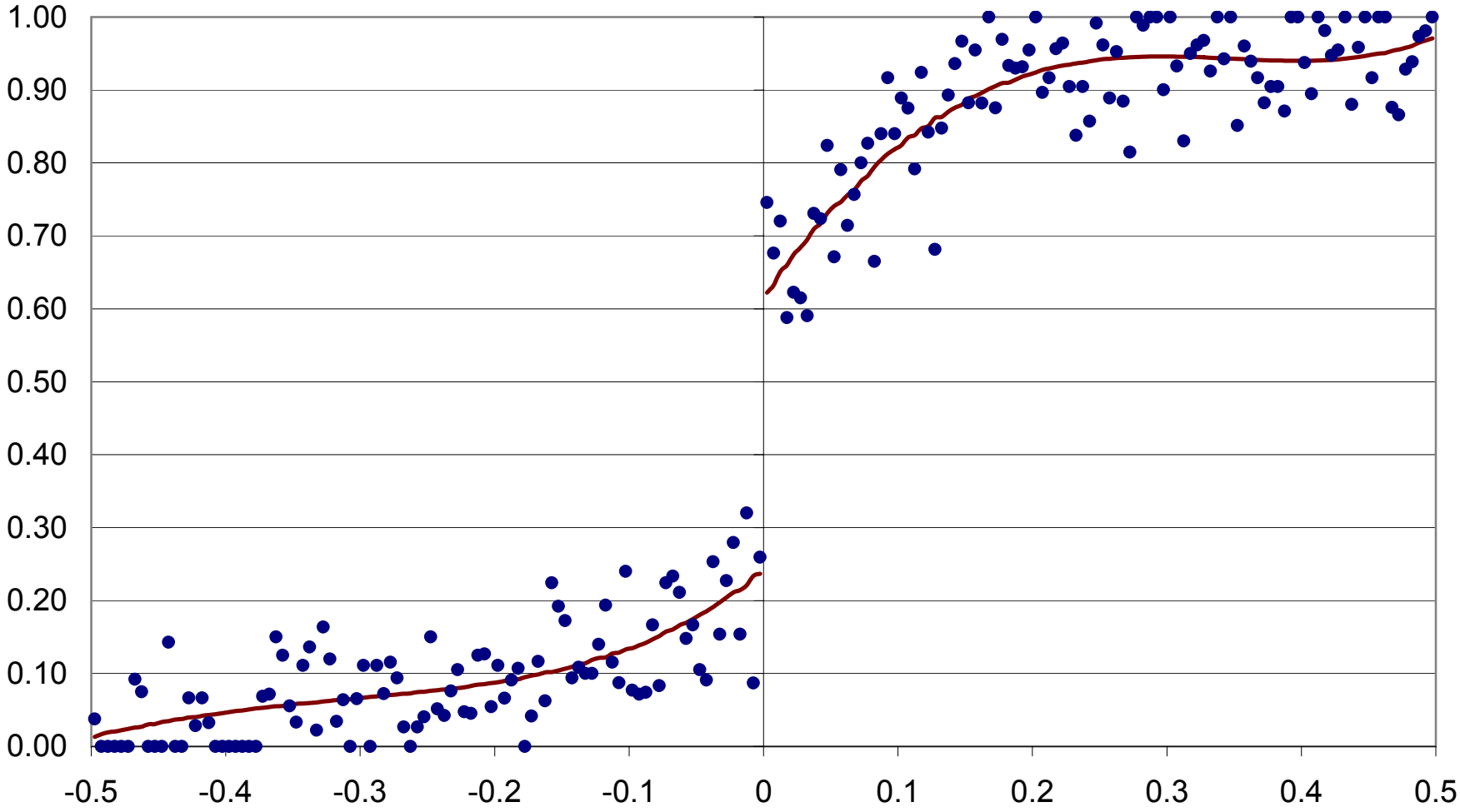


Figure 8: Density of the forcing variable (vote share in previous election)

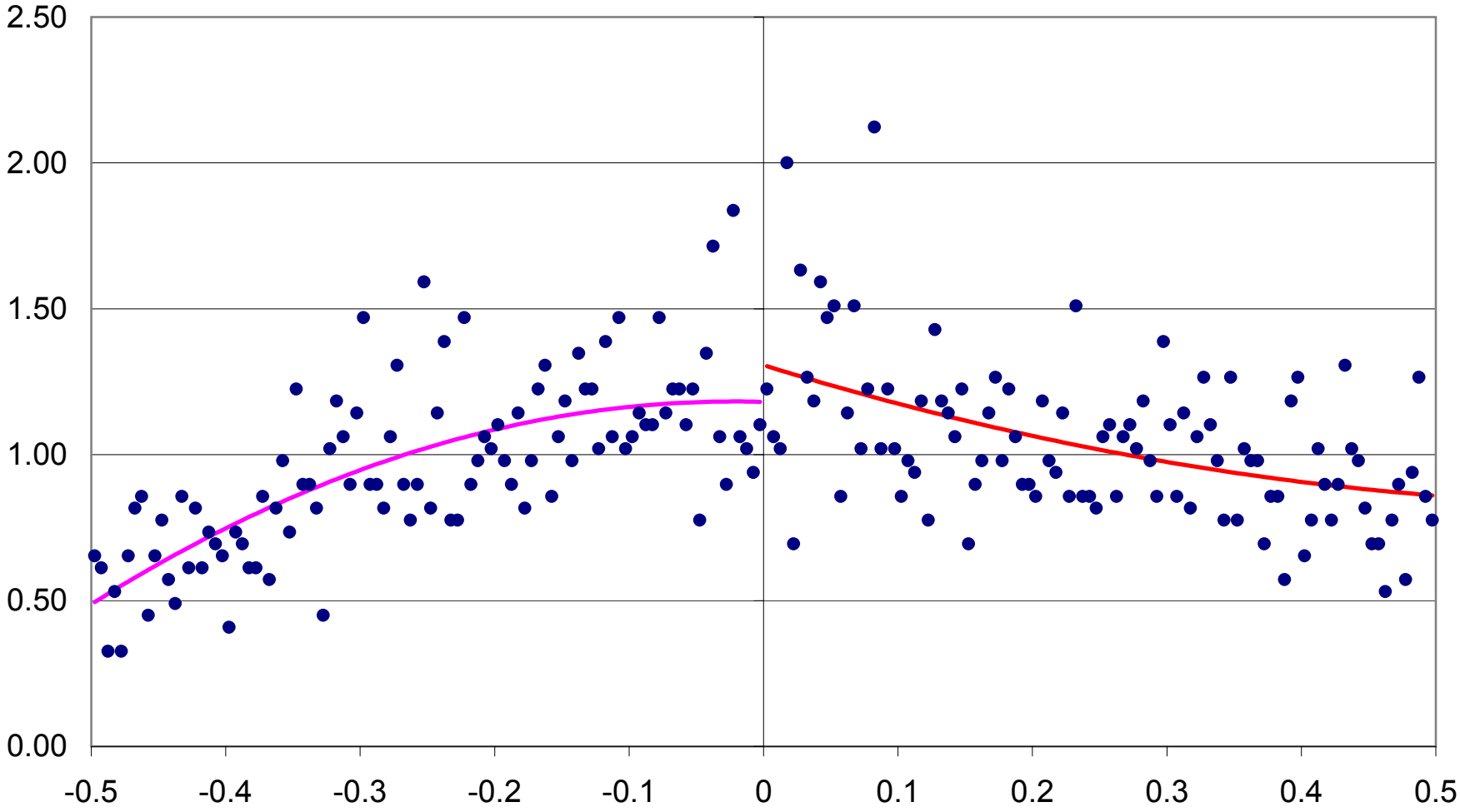
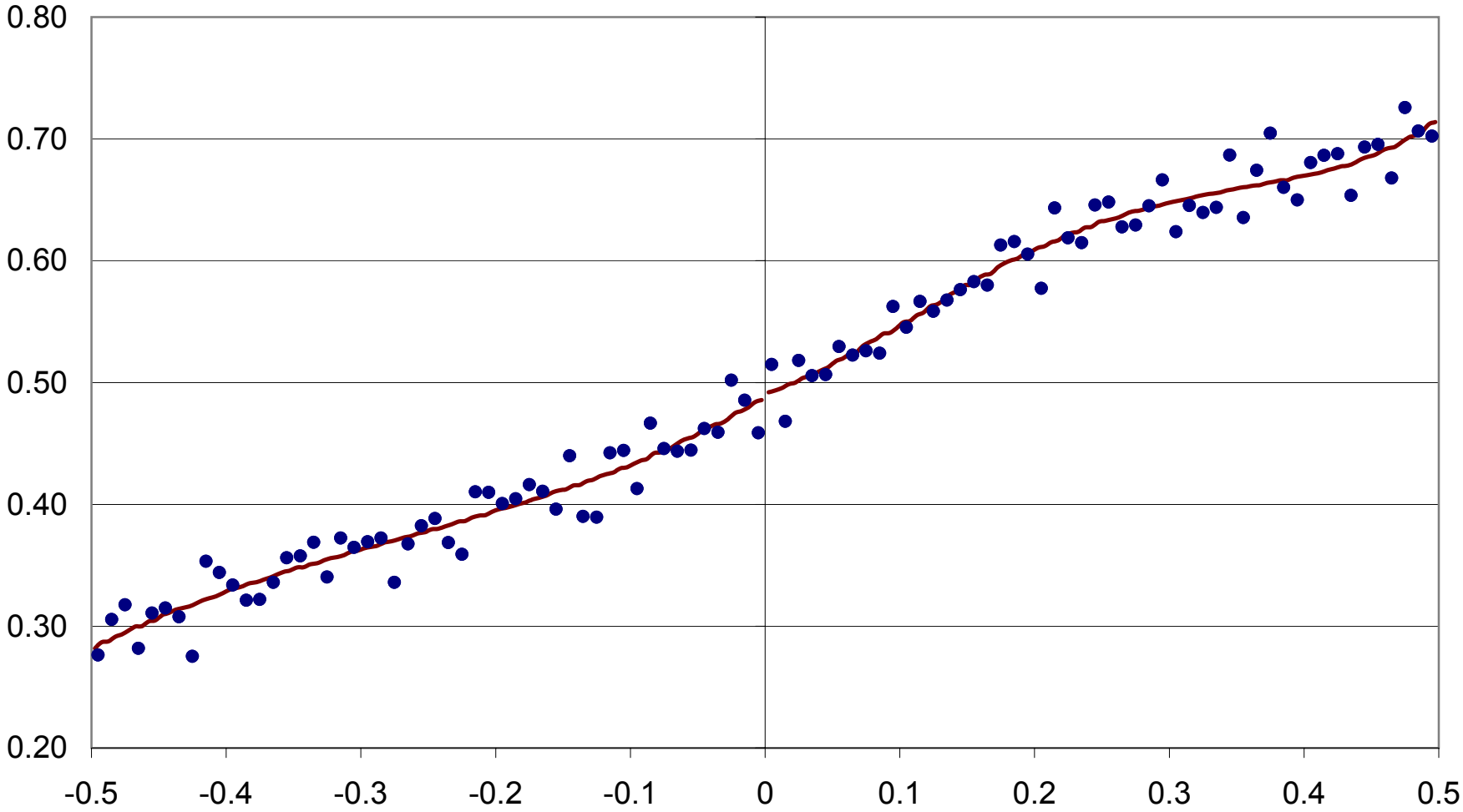


Figure 9: Discontinuity in baseline covariate (share of vote in prior election)



Grupo de Controle Sintético (Abadie e Gardeazabal, *AER* 2003)

- Muitas vezes interesse recai sobre efeitos de eventos ou intervenções que acontecem em níveis agregados

- Exemplos:
  - efeito de salário mínimo sobre emprego: compara-se grupo afetado (eg, estado nos EUA) com um que não tenha sido afetado pela intervenção;
  - efeito de progressão continuada: comparam-se escolas da rede estadual de SP com de outras redes estaduais não afetadas.

- Em geral, para se estimar efeitos de eventos agregados, acompanhamos a evolução temporal da unidade afetada pela intervenção e comparamos essa evolução com a de uma outra unidade que não tenha sido afetada.

- Por exemplo, vemos o que aconteceu com salários e desemprego médio antes e depois da intervenção para tratados e controles (diff-in-diff).

- Como escolher grupos de comparação? No caso do progressão continuada em SP, qual UF usar como comparação? RJ? O fato de ser estado contíguo não pode ter efeito nos municípios fronteiriços?

- Descrição do Método
  - Suponha que haja J+1 regiões e que 1a região foi afetada pelo tratamento (intervenção)
  - Existem disponíveis J regiões de controle. Seja $W = [w_2, ..., w_{J+1}]^{\mathsf{T}}$, onde $w_j \geq 0$ for $j = 2, ..., J+1$ e $w_2 + ... + w_{J+1} = 1$. Cada valor de W representa um controle potencial.
  - Seja X_1 um vetor $k \times 1$ de características pré–tratamento. De mode similar, X_0 é uma matriz $k \times J$ que contém as mesmas variáveis nas regiões não afetadas.
  - O vetor W^* é escolhido como o argumento que minimiza $\|X_1 - X_0 W\|$ sujeito a $w_j \geq 0$ for $j = 2, ..., J+1$ e $w_2 + ... + w_{J+1} = 1$.

– Em geral, $\|A\|_V = \sqrt{A^\intercal V A}$ onde V é matriz simétrica, positiva e semidefinida.

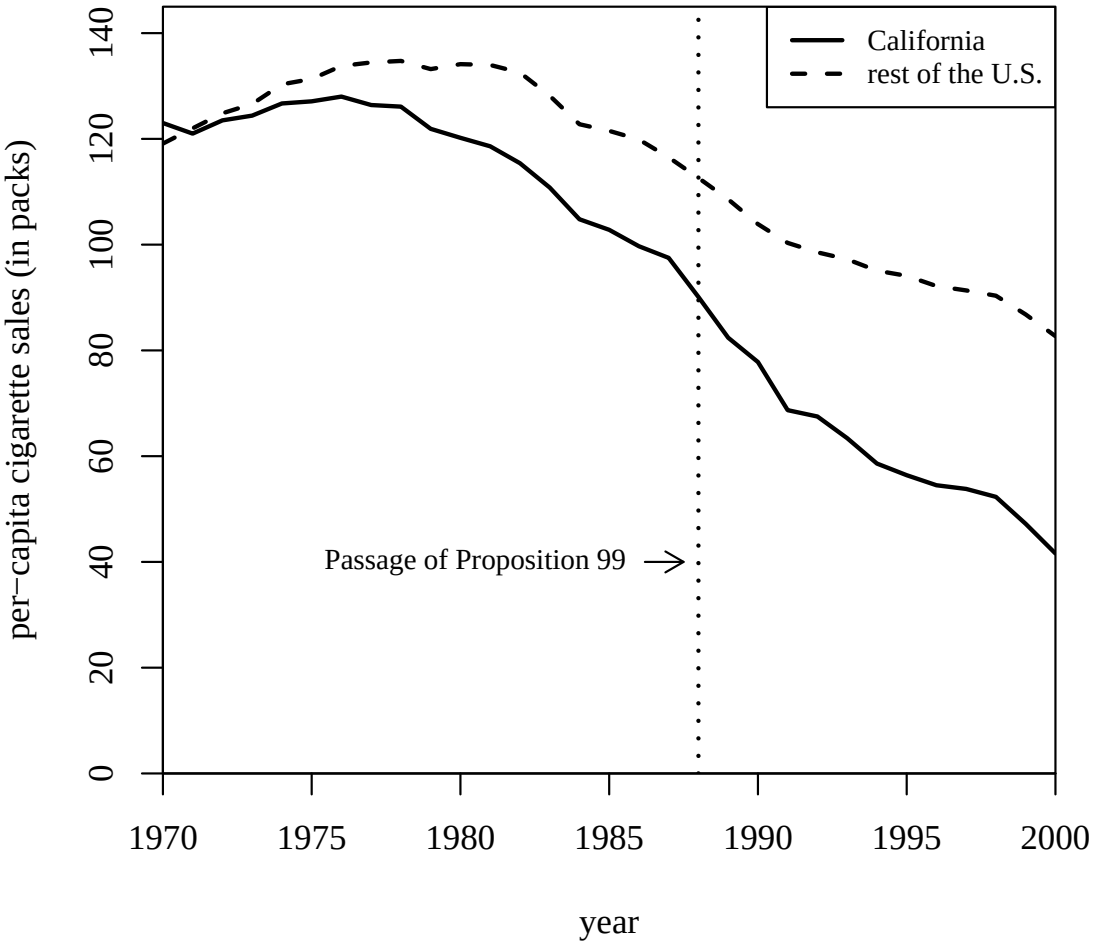
- Vantagens
  - Como controle, usamos a combinação convexa das unidades de comparação que mais se assemelham ao valor das características observáveis do grupo afetado antes do tratamento.
  - Evita-se extrapolação
  - Força pesquisador a demonstrar afinidades entre unidades afetadas e não-afetadas usando caracterísitcas quantificáveis.
  - Pode acomodar presença de fatores não-observáveis.
  - Inferência válida e independente do número de unidades disponíveis para comparação.

- Exemplo: California's Proposition 99, a qual aumentou imposto sobre cigarros em 25 cents/pack.

- Ver em "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program" de Alberto Abadie, A. Diamond and J. Hainmueller (2007).

California's Proposition 99

- In 1988, California was first to pass comprehensive tobacco control legislation (Proposition 99)
  - increased cigarette tax by 25 cents/pack
  - earmarked tax revenues to health and anti-smoking budgets
  - funded anti-smoking media campaigns
  - spurred clean-air ordinances throughout the state
  - produced more than $100 million per year in anti-tobacco projects
- What was the effect of Proposition 99 on tobacco consumption in California?

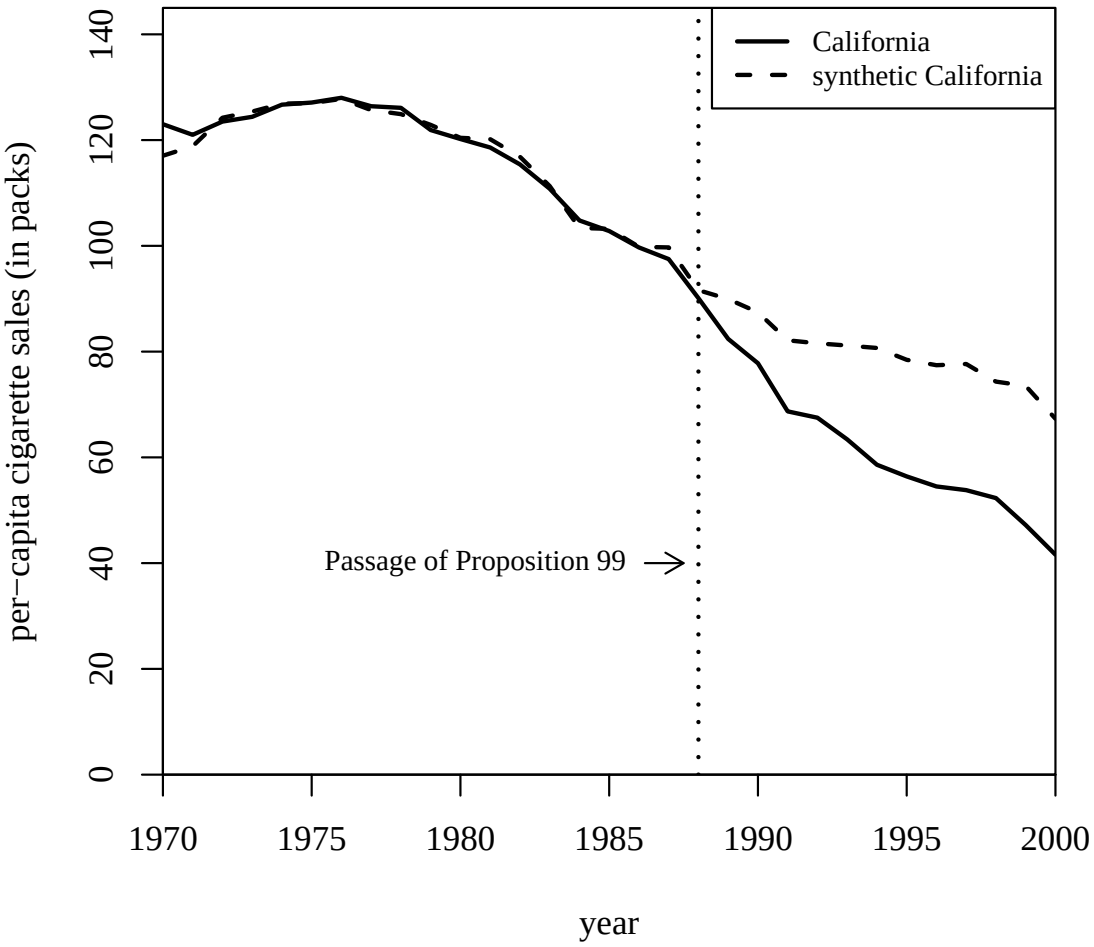


California and the Rest of the U.S.

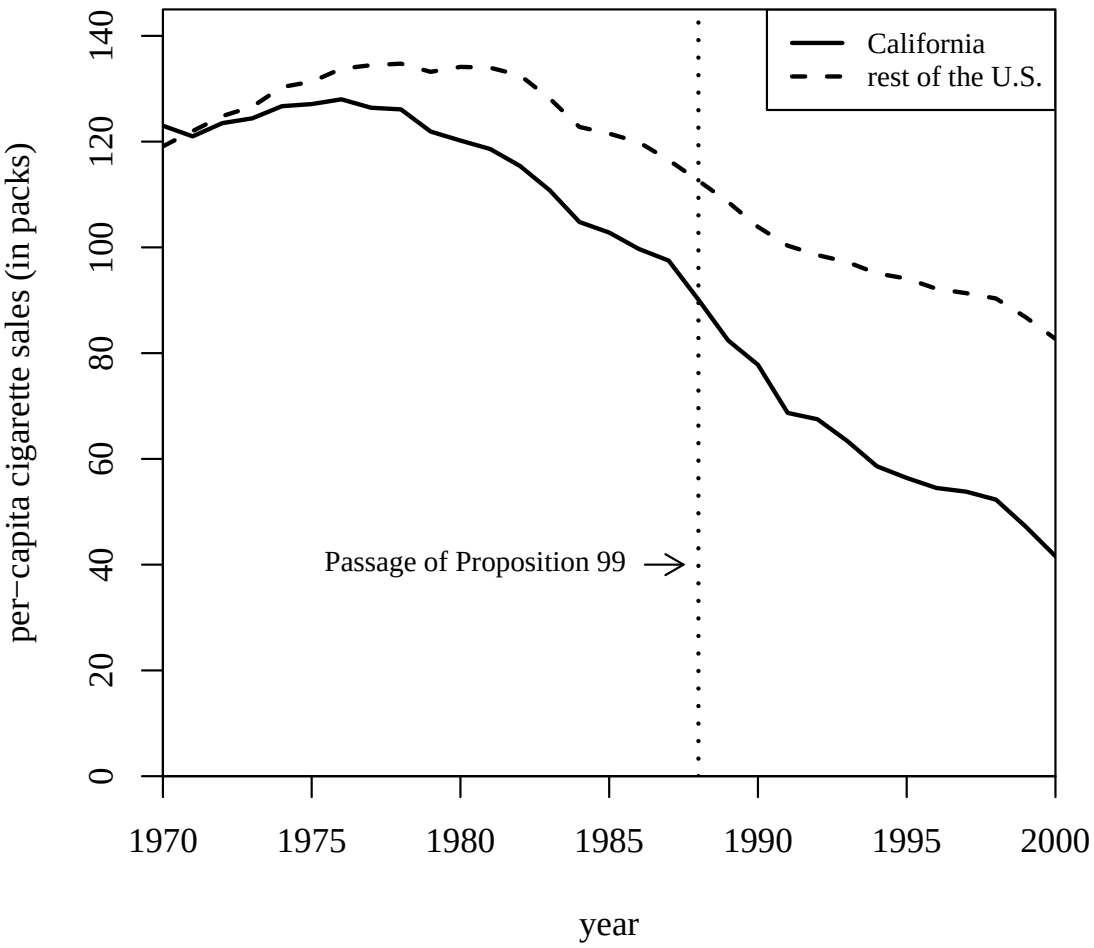
Smoking Prevalence Predictors Means

Variables	California		Average of
	Real	Synthetic	38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

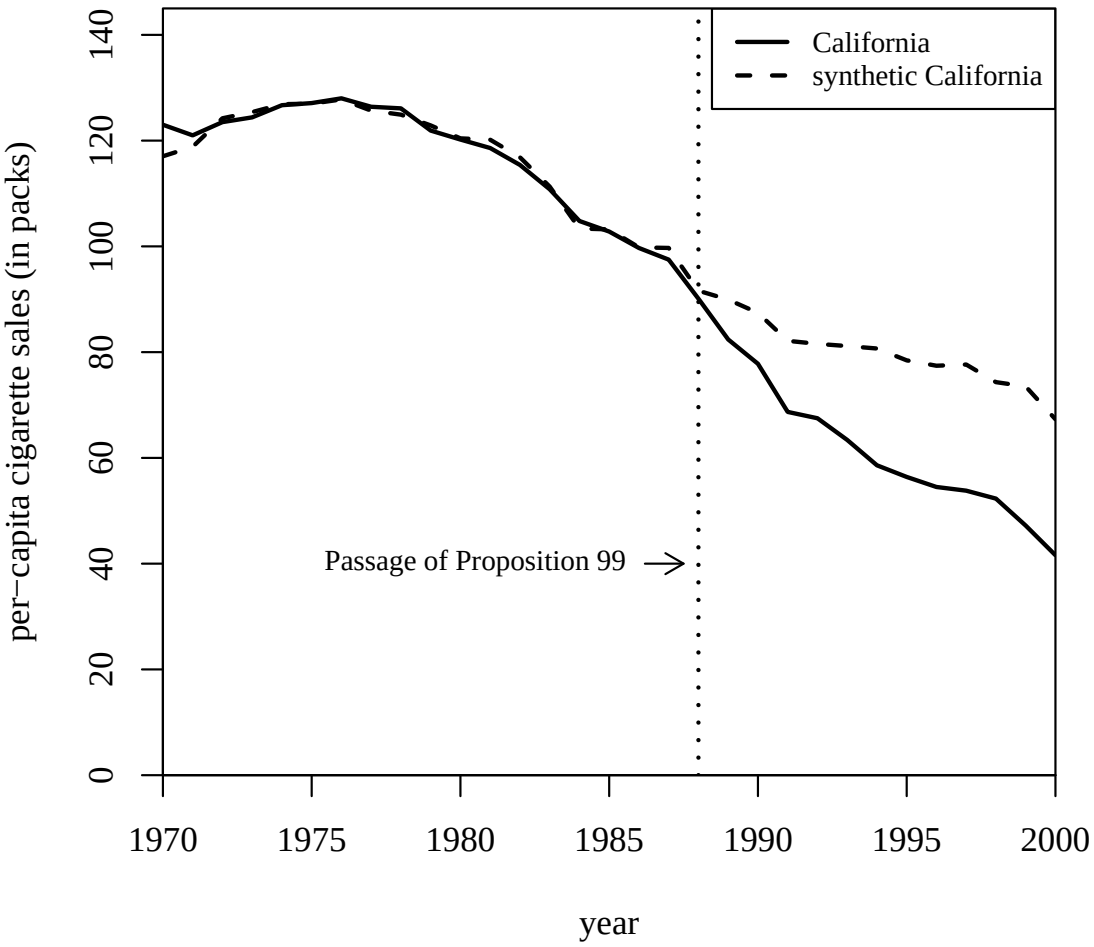
Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).



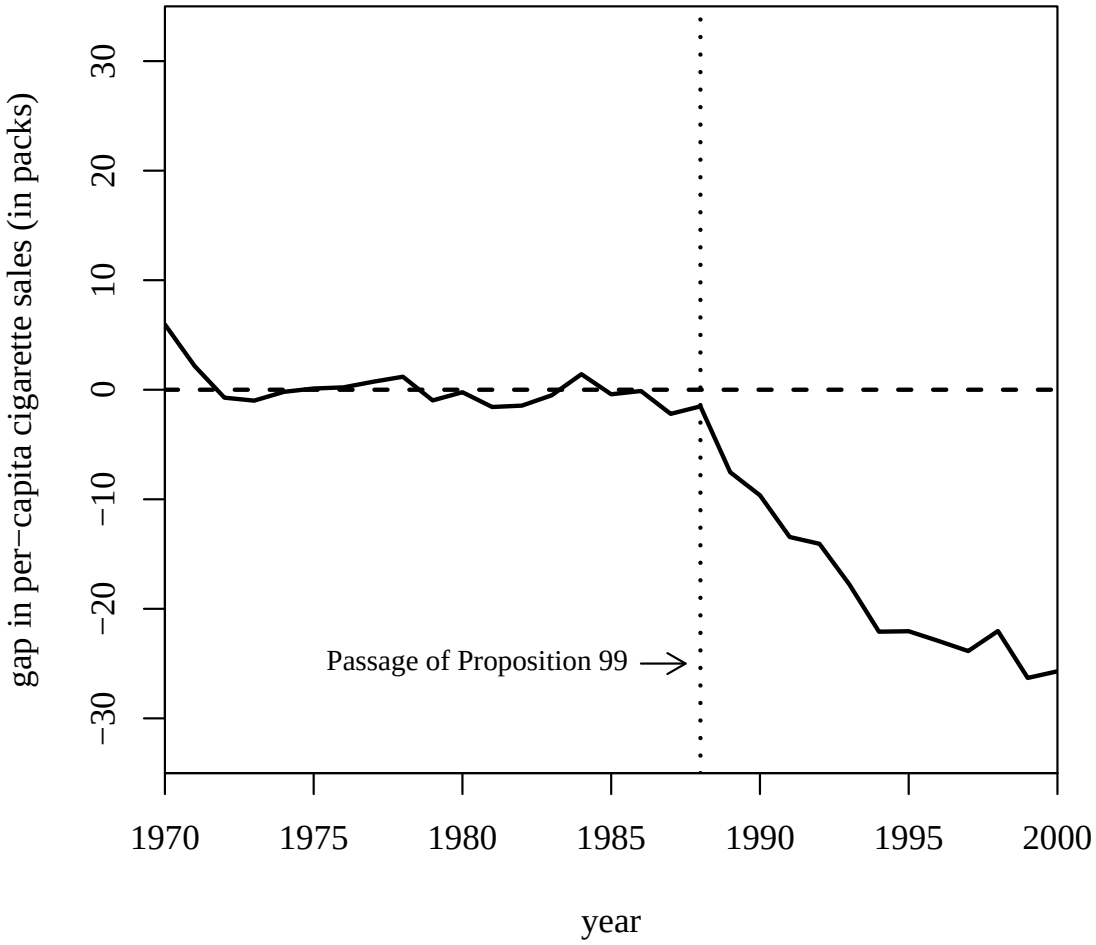
California and synthetic California



California and the Rest of the U.S.



California and synthetic California



Smoking Gap Between California and synthetic California