

# Regression Discontinuity

EC 607, Set 9

---

Edward Rubin  
Spring 2020

# Prologue

# Schedule

## Last time

- Introduction to selection-on-unobservables designs
- Instrumental variables (IV) and two-stage least squares (2SLS)

## Today

Regression discontinuity <sup>†</sup>

## Upcoming

- Problem set; office hours Friday?
- Project

<sup>†</sup> These notes largely follow notes from [Michael Anderson](#), [Imbens and Lemieux \(2008\)](#), and notes from [Teppei Yamamoto](#).

# Regression discontinuity

# Regression discontinuity

## Setup

We're still in the game of estimating the effect of a potentially endogenous treatment  $D_i$  on an outcome  $Y_i$ .

**Regression discontinuity (RD)** offers a particularly clear/clean research design based upon an arbitrary threshold (the *discontinuity*).

That said, most RDs boil down to an implementation of IV.

In addition, while RD is all the rage in modern applied econometrics, **Thistlewaite and Campbell** wrote about it back in 1960.

# Regression discontinuity

## Our framework

Back to our potential-outcome framework.

We want to know the effect of  $D_i$  on  $Y_i$ .

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$$

**New:** Suppose  $D_i$  is determined<sup>†</sup> by whether some variable  $X_i$  crosses a threshold  $c$  (the discontinuity).

The variable  $X_i$  need not be randomly assigned—we will assume it is not (*i.e.*,  $X_i$  correlates with  $Y_{0i}$  and  $Y_{1i}$ ).

We will assume that  $Y_{0i}$  and  $Y_{1i}$  vary smoothly in  $X_i$ .

<sup>†</sup> At least in part.

# Regression discontinuity

## Examples

We often apply regression-discontinuity designs in settings with some arbitrary threshold embedded within some bureaucratic decision.

- An elector candidate wins if her vote share exceeds her competitors.
- Election runoffs are triggered if "winner" is below 50%.
- Antidiscrimination laws only apply to firms with >15 employees.
- Prisoners are eligible for early parole if some score exceeds a threshold.
- An individual is eligible for Medicare if her age is at least 65.
- You get a ticket if your speed exceeds the speed limit.
- Fifteen-percent discount at Sizzler if your age exceeds 60.
- Counties with  $PM_{2.5} > 35 \mu\text{g}/\text{m}^3$  are *out of attainment*.

In some cases, "treatment" is definite once we exceed the threshold.

# Regression discontinuity

## Sharp vs. fuzzy

We distinguish RDs by how strong/definitive the threshold is.

In **sharp RDs**, individuals move from control to treatment when their  $\mathbf{X}_i$  passes our threshold  $c$ , *i.e.*,  $\mathbf{D}_i$  switches from 0 to 1 when  $\mathbf{X}_i$  moves across  $c$ .

*E.g.*, a politician wins an election when the difference between her vote share and her competitor's vote share exceeds zero.

In **fuzzy RDs**, the *probability of treatment*  $\Pr(\mathbf{D}_i = 1)$  discontinuously jumps at the threshold  $c$ , but it does not move from 0 to 1.

*E.g.*, crossing some GRE threshold discontinuously increases your chances of getting into some grad schools (but doesn't guarantee admittance).

Sharp RDs

# Sharp RDs

## Setup

With **sharp regression discontinuity**, the probability of treatment changes from 0 to 1 as  $\mathbf{X}_i$  moves across threshold  $c$ .

Thus, treatment status entirely depends upon whether  $\mathbf{X}_i \geq c$ , *i.e.*,

$$\mathbf{D}_i = \mathbb{I}\{\mathbf{X}_i \geq c\}$$

To estimate the causal effect of  $\mathbf{D}_i$  on  $\mathbf{Y}_i$ , we **compare the mean of  $\mathbf{Y}_i$  just above the threshold to the mean of  $\mathbf{Y}_i$  just below the threshold.**

# Sharp RDs

## More formally

We can write the comparison of means **at the threshold** as

$$\begin{aligned} & \lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x] \\ &= \lim_{x \downarrow c} E[Y_{1i} \mid X_i = x] - \lim_{x \uparrow c} E[Y_{0i} \mid X_i = x] \\ &= \tau_{\text{SRD}} \end{aligned}$$

*Assumption*  $E[Y_{1i} \mid X_i = x]$  and  $E[Y_{0i} \mid X_i = x]$  are continuous in  $x$ .

$$\implies \tau_{\text{SRD}} = E[Y_{1i} - Y_{0i} \mid X_i = c]$$

*i.e.*, Because we don't observe  $Y_{0i}$  for treated individuals, we extrapolate  $E[Y_{0i} \mid X_i = c - \varepsilon]$  to  $E[Y_{0i} \mid X_i = c + \varepsilon]$  for small  $\varepsilon$ .

# Sharp RDs

## Estimation

Thus, we estimate

$$\tau_{\text{SRD}} = \lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x]$$

as the difference between two regression functions estimated "near"  $c$ .

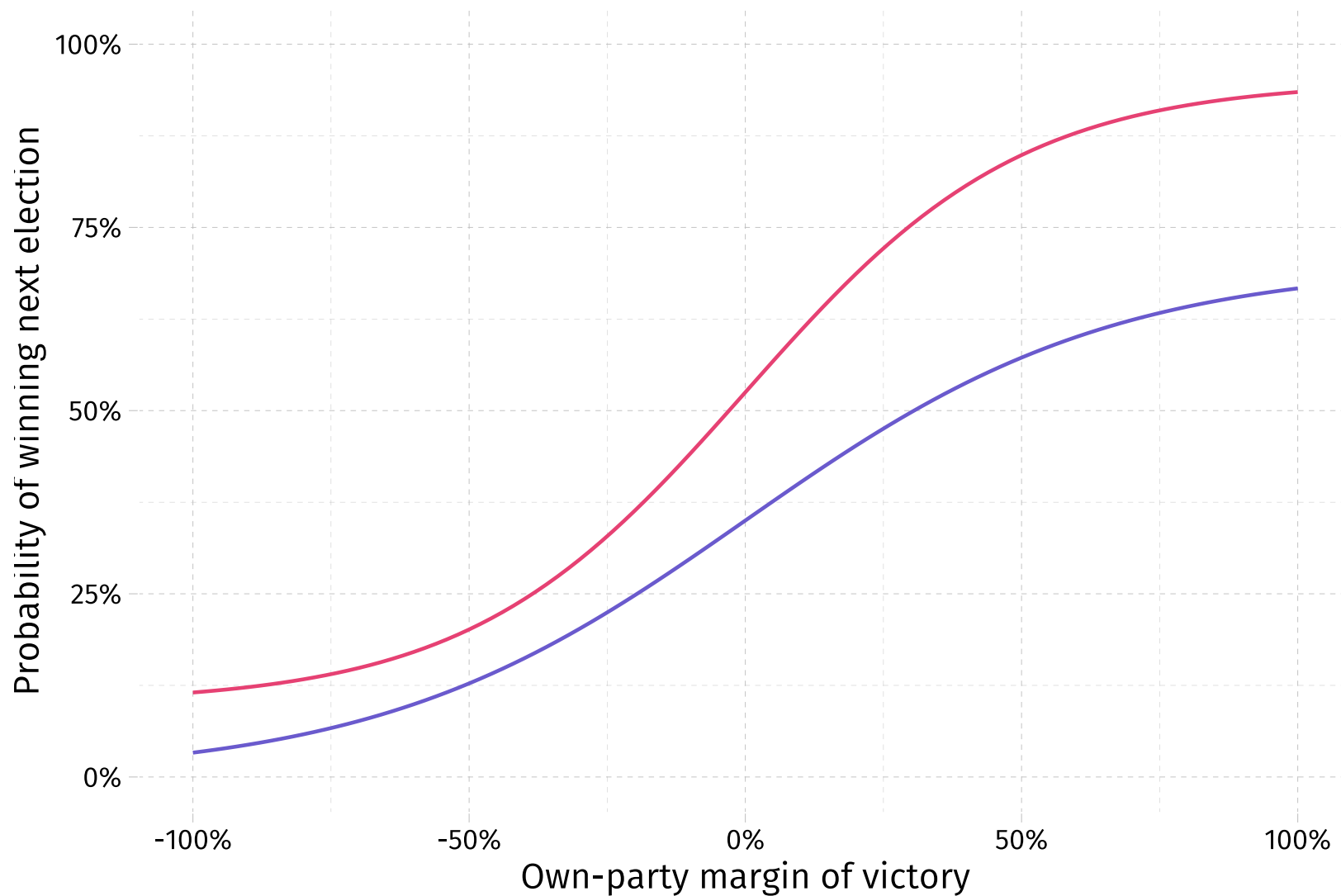
We must stay "near" to  $c$  to minimize the bias from extrapolating  $E[Y_{0i} \mid X_i = c - \varepsilon]$  to  $E[Y_{0i} \mid X_i = c + \varepsilon]$  (and assuming continuity).

*Ex.* Is there effect of a political party winning an election on that party's likelihood of winning the following election?

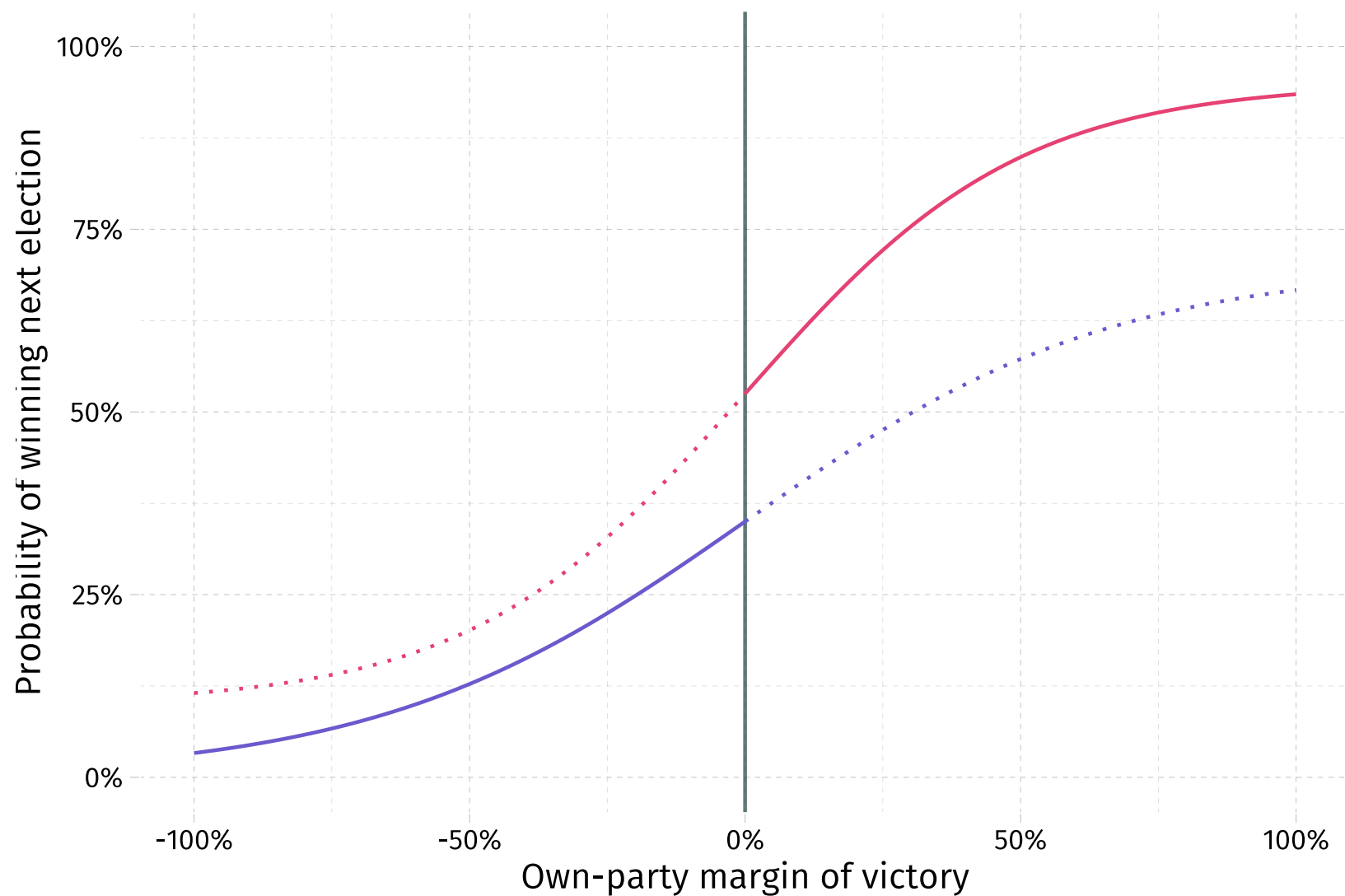
Is there a benefit of incumbency (at the party level)?<sup>†</sup>

<sup>†</sup> Lee (2008) addresses this question via RD. Caughey and Sekhon (2011) discuss RD in this setting.

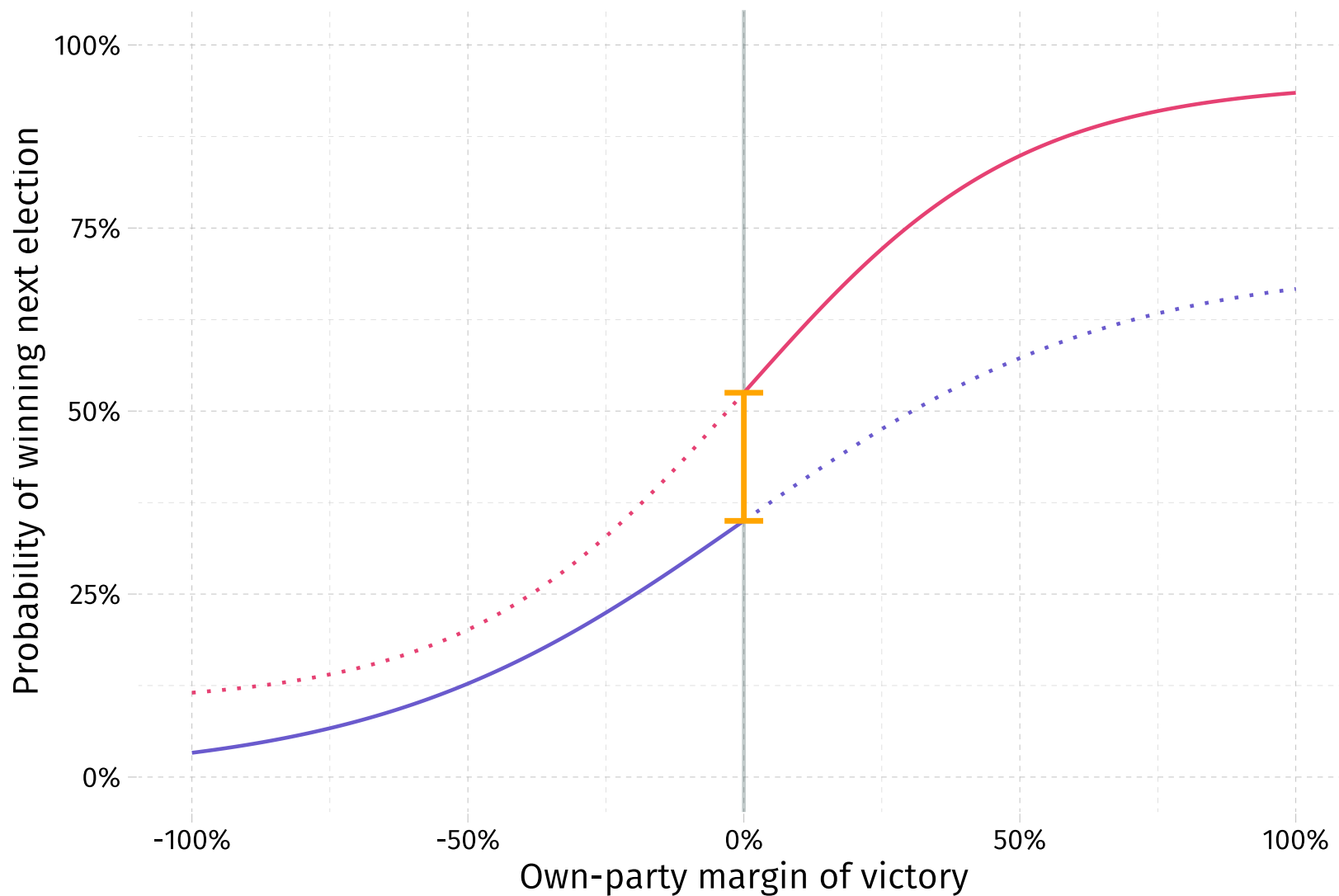
Let's start with  $E[Y_{0i} | X_i]$  and  $E[Y_{1i} | X_i]$ .



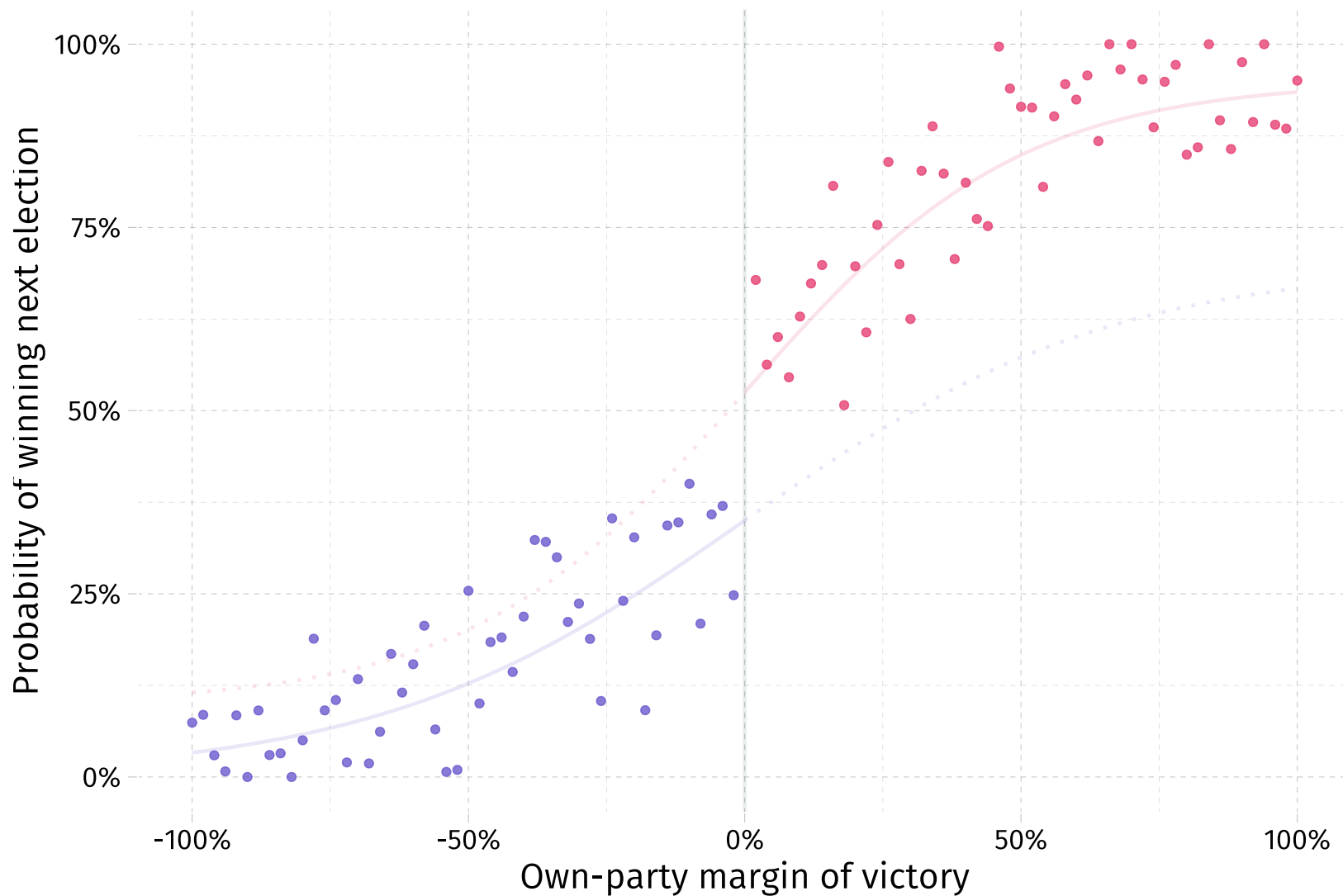
You only win an election if your **margin of victory exceeds zero**.



$E[Y_{1i} | X_i] - E[Y_{0i} | X_i]$  at the discontinuity gives  $\tau_{\text{SRD}}$ .



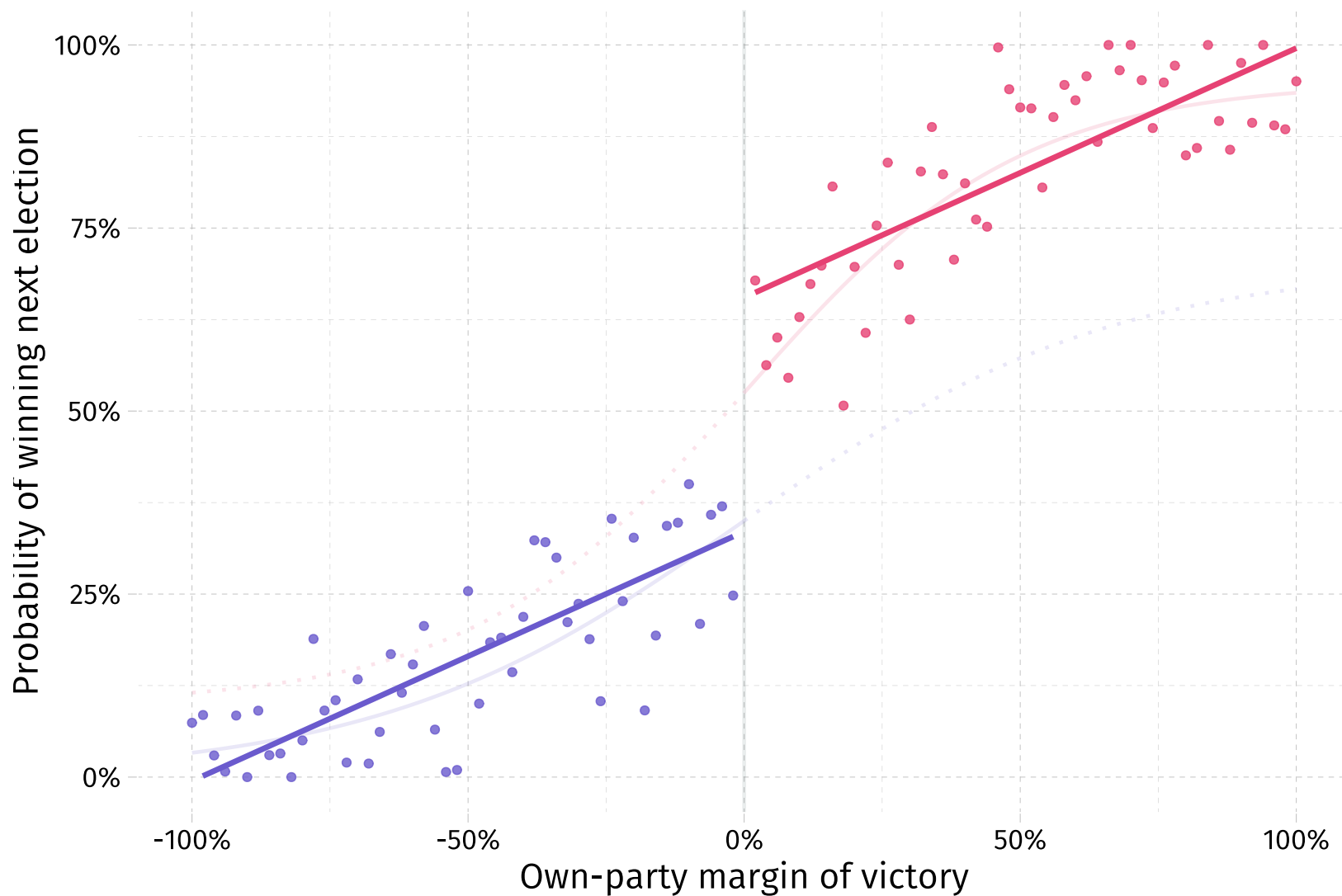
Real data are a bit trickier. We must estimate  $E[Y_{1i} | X_i]$  and  $E[Y_{0i} | X_i]$ .



## Questions

1. How should we estimate  $E[Y_{1i} | X_i]$  and  $E[Y_{0i} | X_i]$ ?
2. How much data should we use—*i.e.*, what is the right **bandwidth** size?

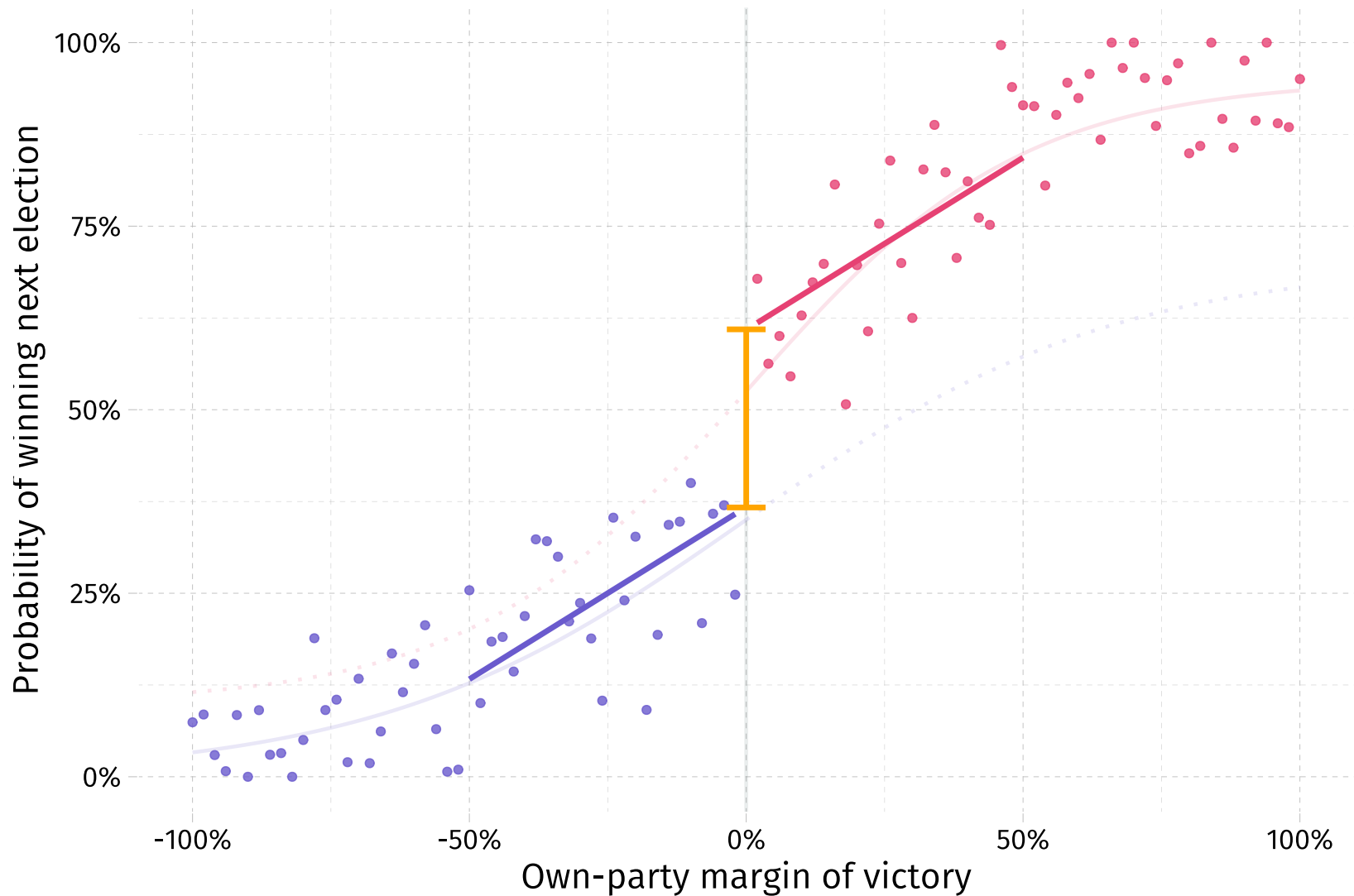
## Option 1a Linear regression with constant slopes (and all data)



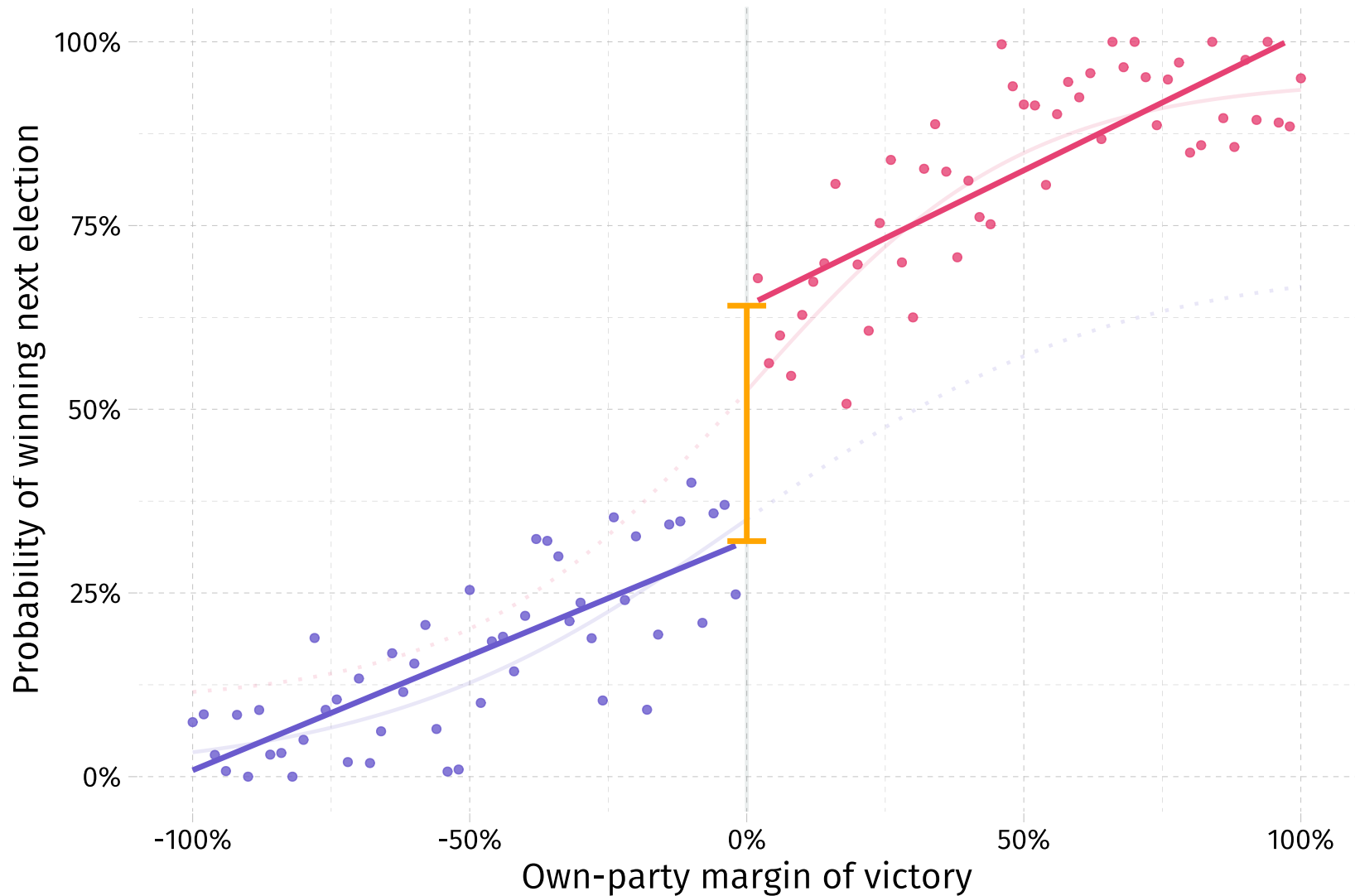
*Option 1a* Linear regression with constant slopes (and all data)



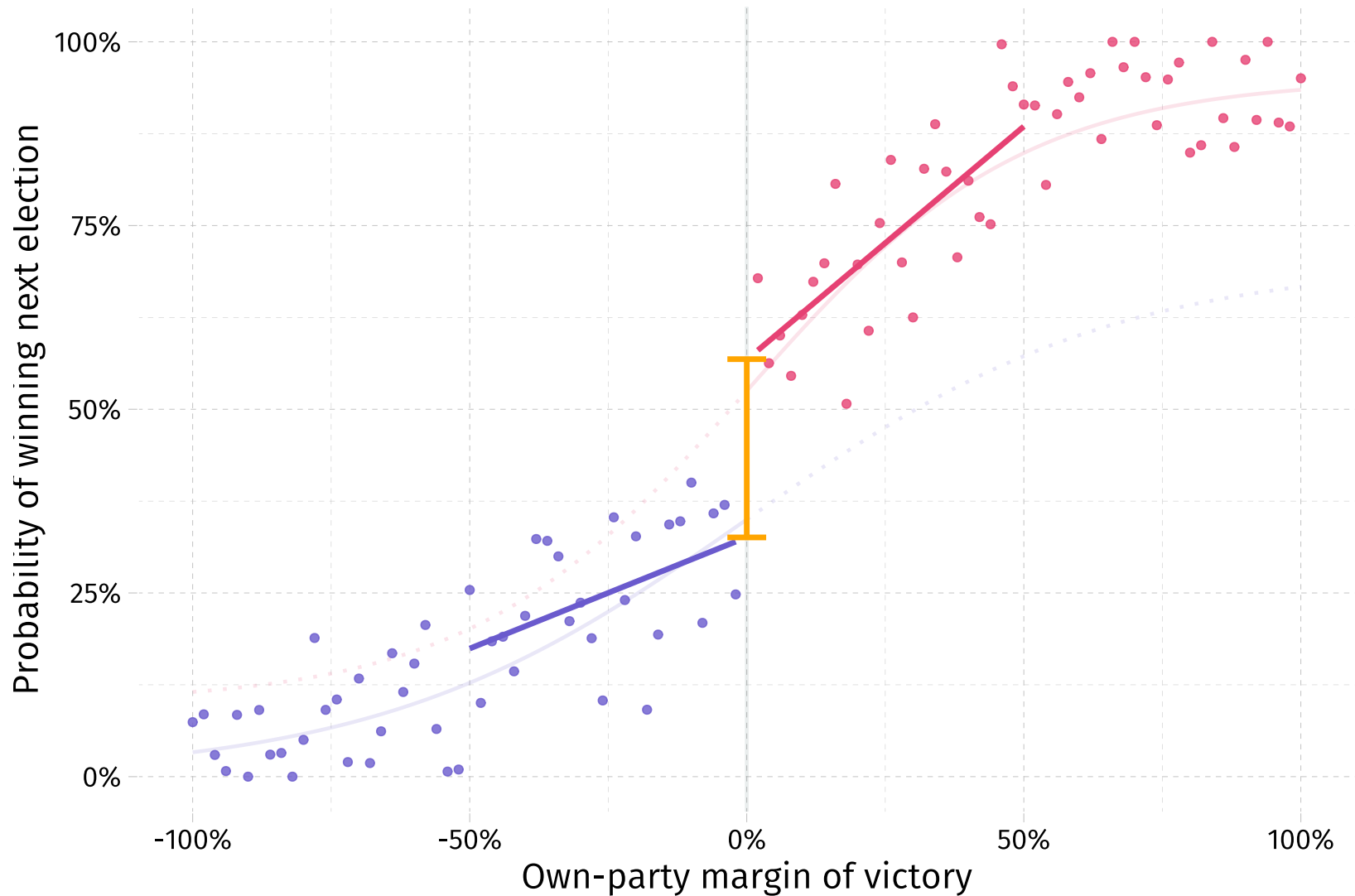
*Option 1b* Linear regression with constant slopes; limited to  $\pm 50\%$ .



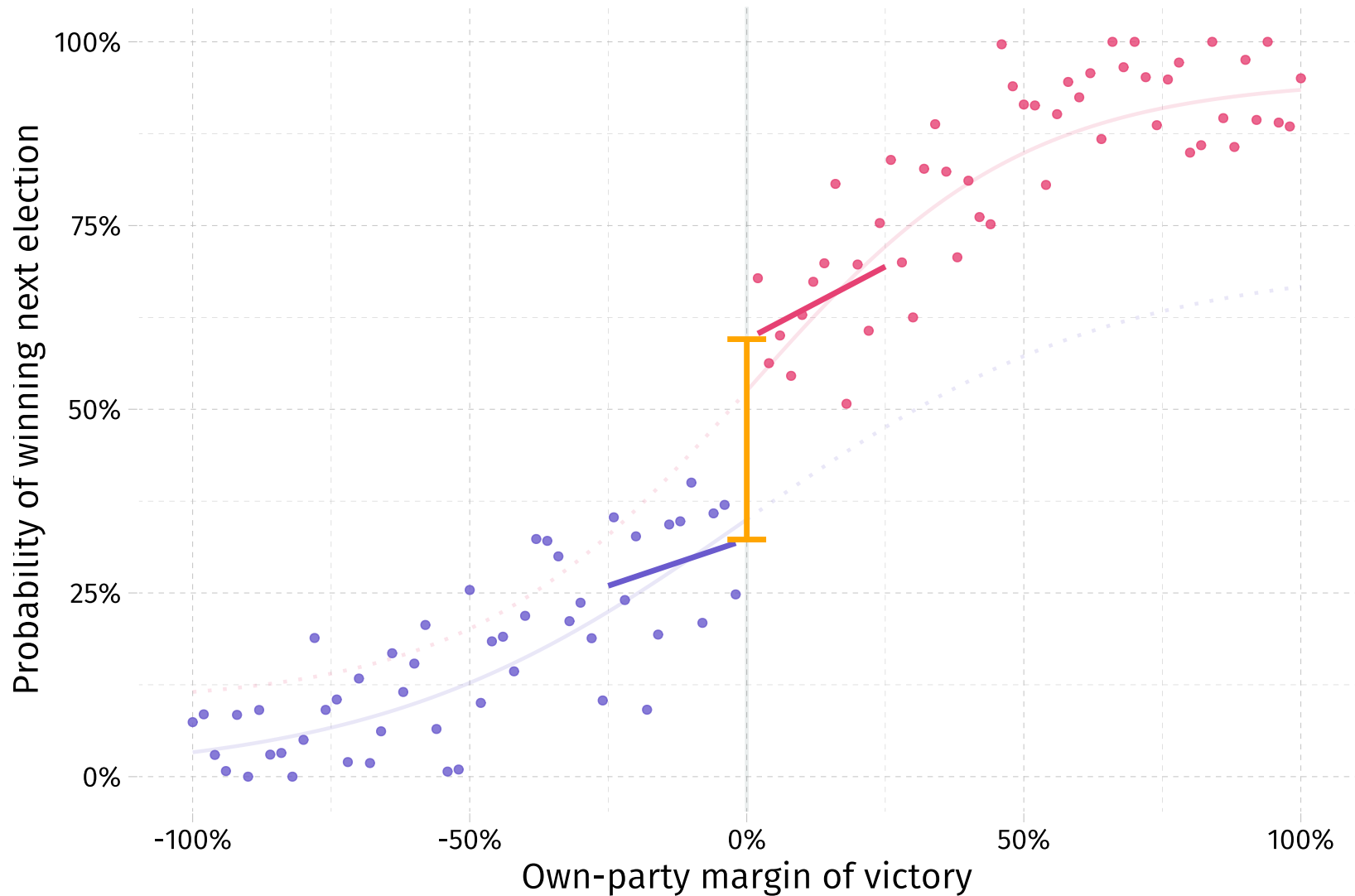
## Option 2a Linear regression with differing slopes (and all data)



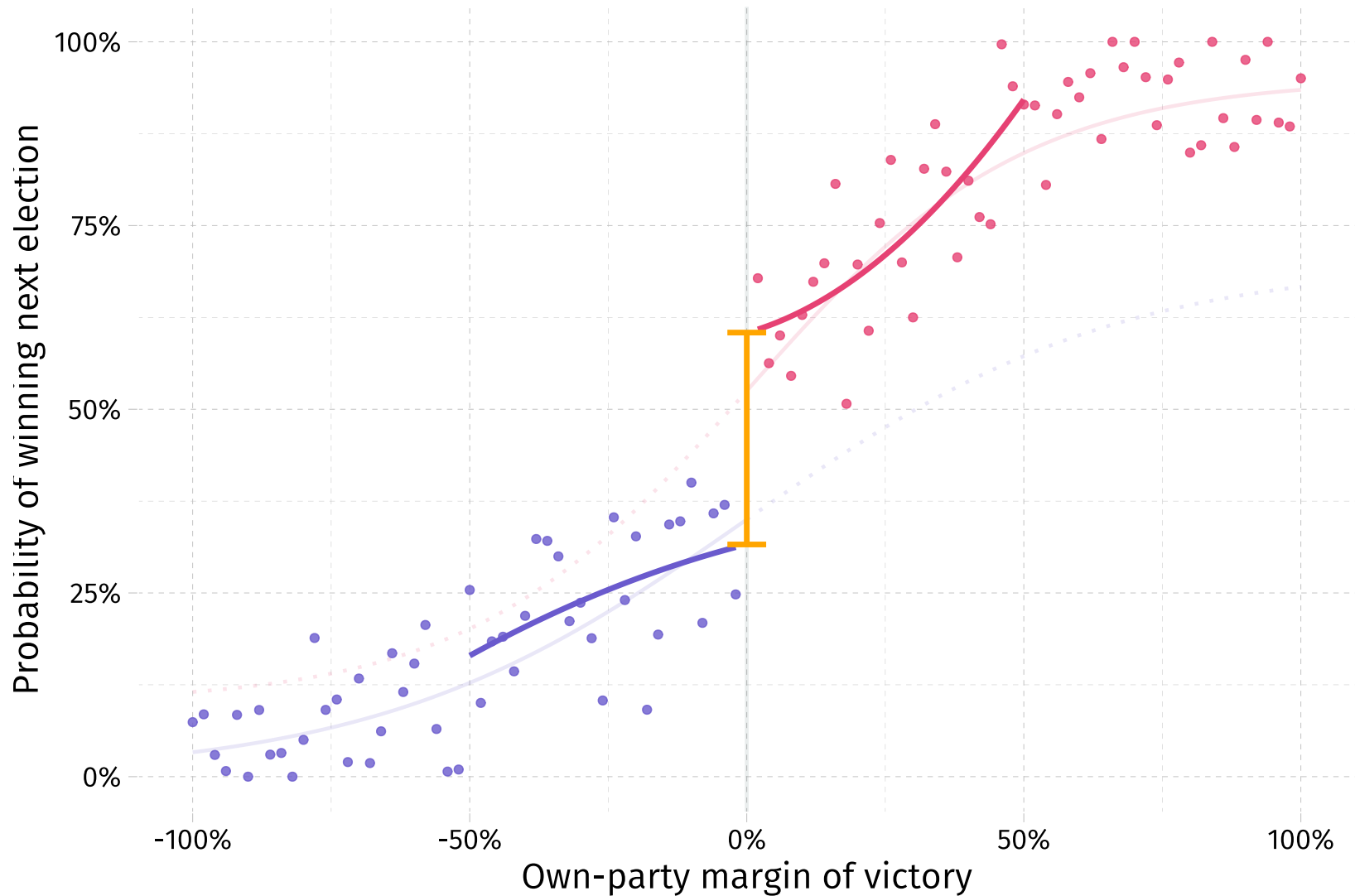
*Option 2b* Linear regression with differing slopes; limited to  $\pm 50\%$ .



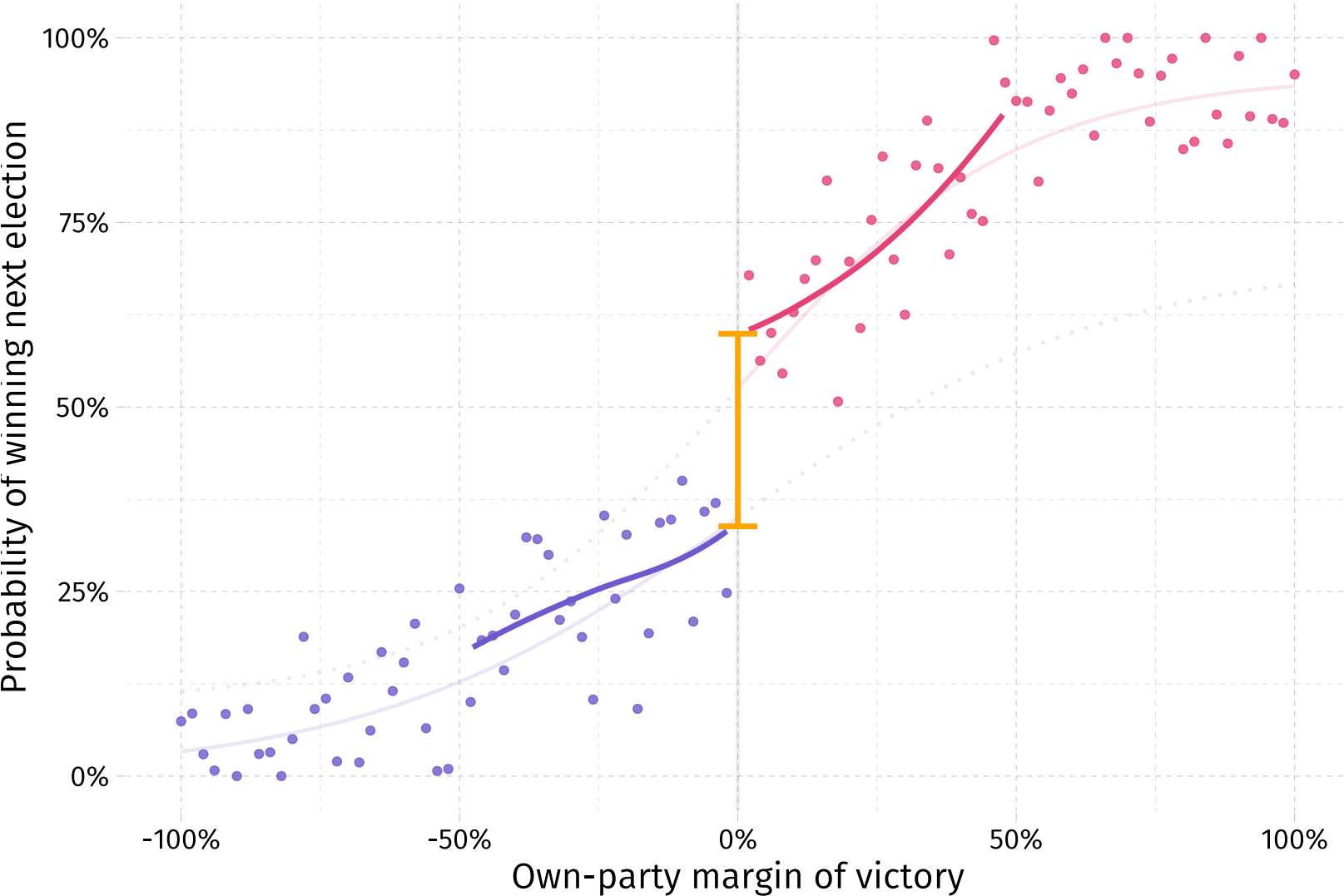
*Option 2c* Linear regression with differing slopes; limited to  $\pm 25\%$ .



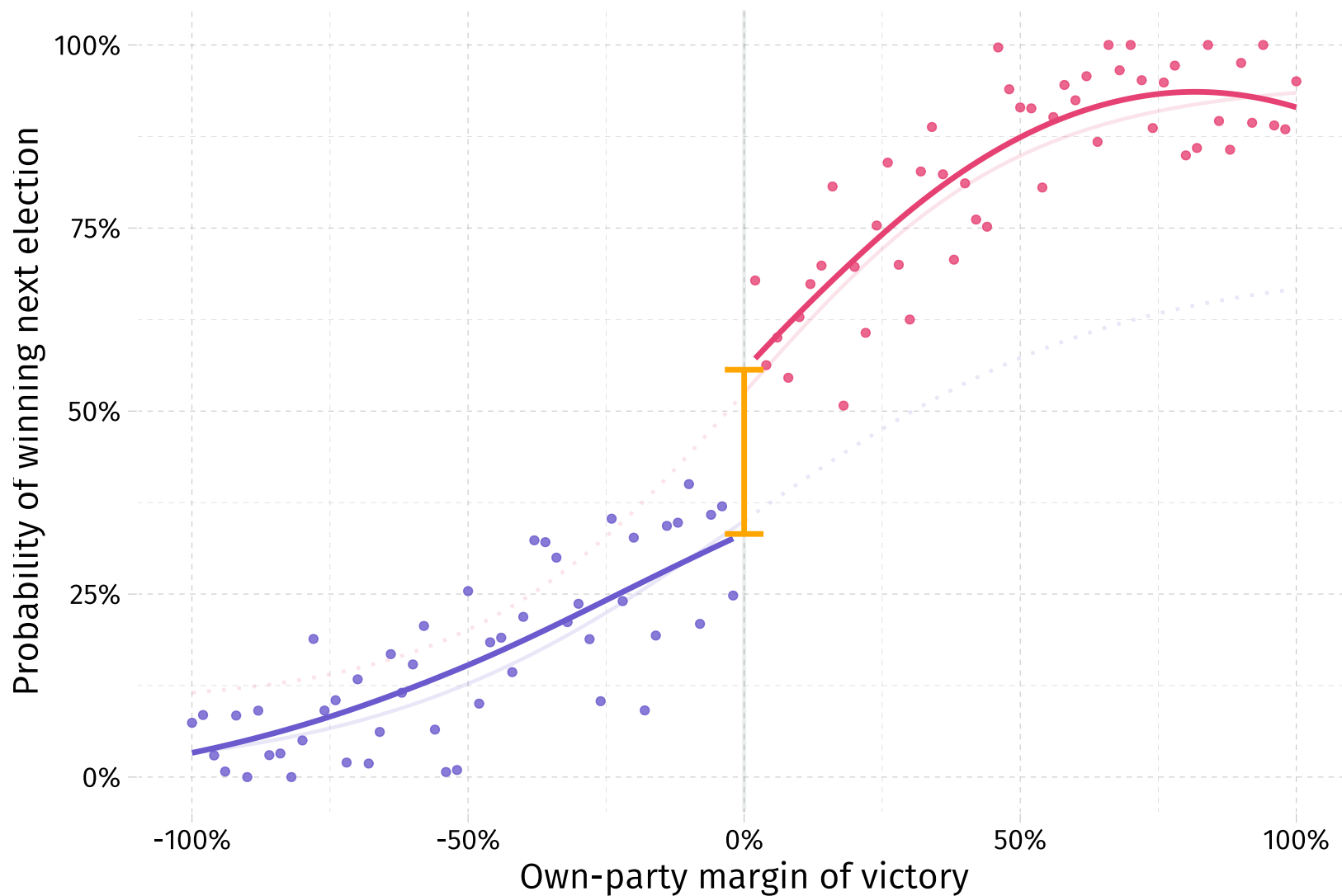
Option 3 Differing quadratic regressions (limited to  $\pm 50\%$ ).



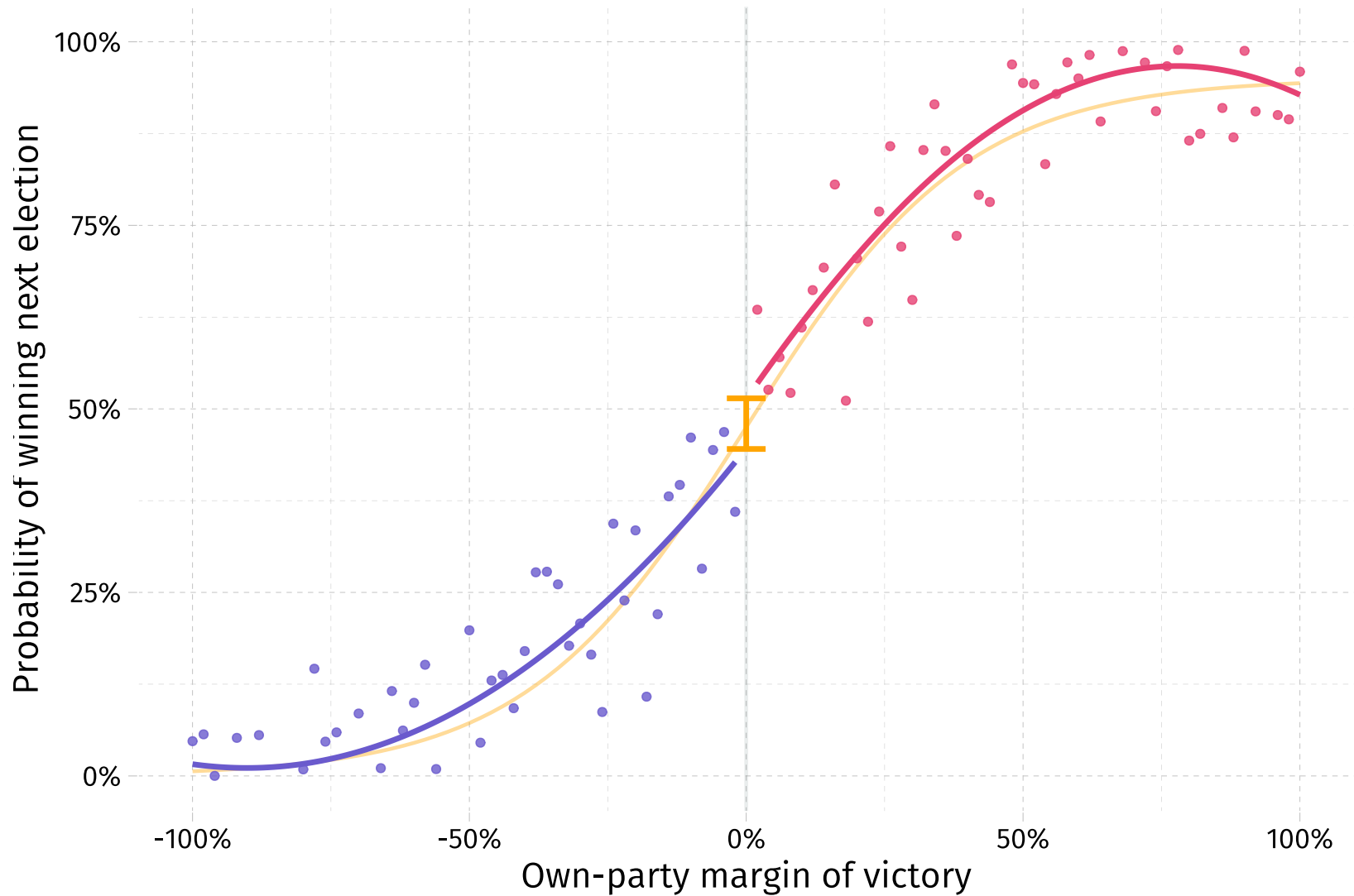
Option 4a Differing local (LOESS) regressions (limited to +/- 50%).



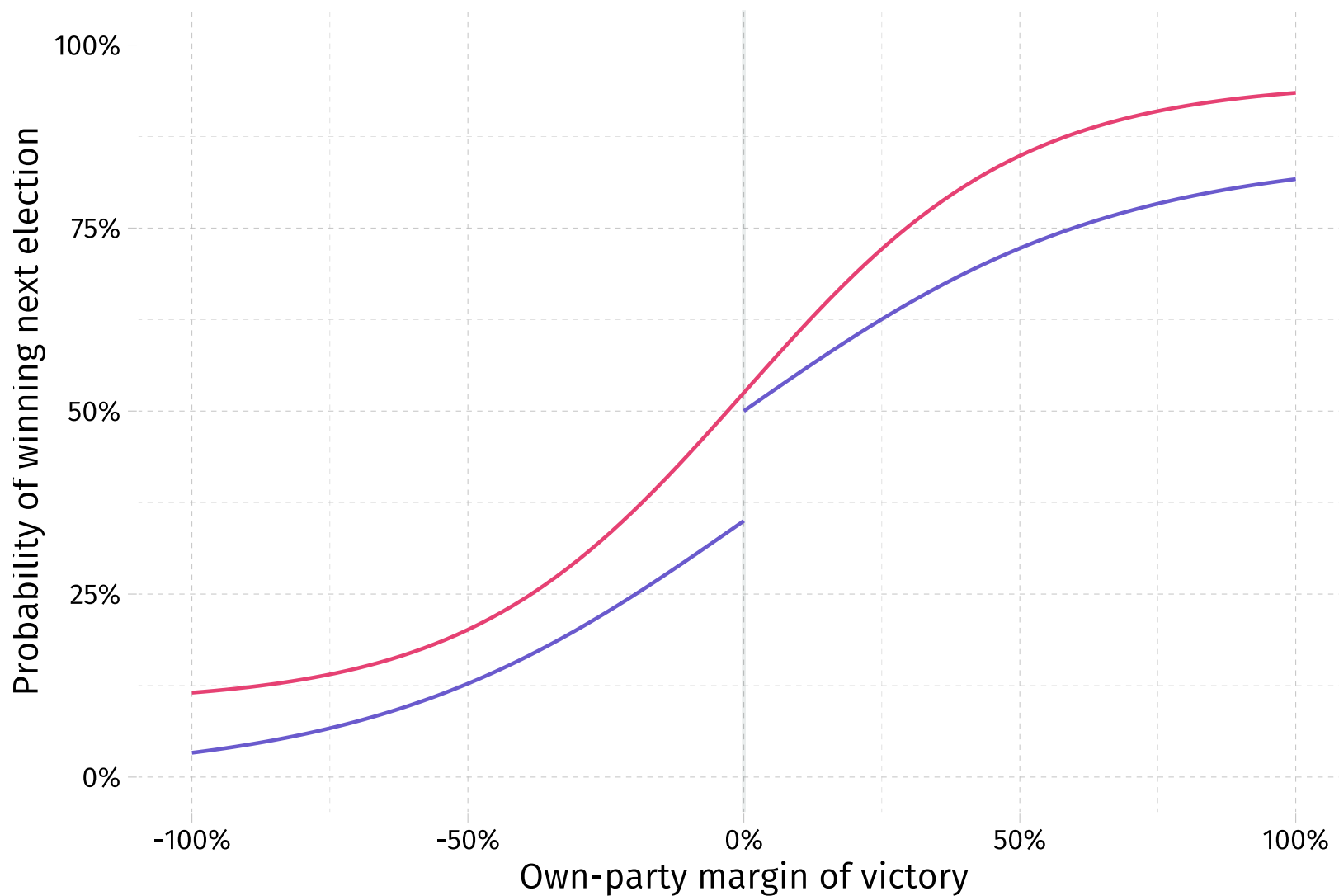
*Option 4b* Differing local (LOESS) regressions (all data).



*Note* Functional form can be very important.



The continuity of  $E[Y_{0i} | X_i = x]$  (in  $x$ ) is also very important. No sorting.



# Sharp RDs

## In practice

Gelman and Imbens (2018) on functional form:

We argue that controlling for global high-order polynomials in regression discontinuity analysis is a flawed approach with three major problems: it leads to noisy estimates, sensitivity to the degree of the polynomial, and poor coverage of confidence intervals. We recommend researchers instead use estimators based on local linear or quadratic polynomials or other smooth functions.

See Imbens and Kalyanaraman (2012) for optimal bandwidth selection.

# Sharp RDs

## Estimation

1. **Trim data** to a reasonable window around the threshold  $c$ .
2. **Recode  $X_i$**  (the "forcing variable") as deviation from  $c$ , *i.e.*,  $\tilde{X}_i = X_i - c$ 
  - $\tilde{X}_i = 0$  if  $X_i = c$
  - $\tilde{X}_i < 0$  if  $X_i < c$  and thus  $D_i = 0$
  - $\tilde{X}_i > 0$  if  $X_i > c$  and thus  $D_i = 1$
3. Determine a model to **estimate**  $E[Y_i | \tilde{X}_i]$  for  $\tilde{X}_i$  above and below 0
  - Linear with common slopes for  $E[Y_i | \tilde{X}_i < 0]$  and  $E[Y_i | \tilde{X}_i > 0]$
  - Linear/quadratic/polynomial with differing slopes
  - LOESS, kernel regression, *etc.*

# Sharp RDs

## Estimation: Linear, common slope

### *Assumptions*

1.  $E[Y_{0i} | X_i = x]$  is linear in  $x$ , i.e.,  $E[Y_{0i} | X_i] = \alpha + \beta X_i$
2. Treatment effect does not depend upon  $X_i$ , i.e.,  $E[Y_{1i} - Y_{0i} | X_i] = \tau$

where (1) comes from linearity and (2) comes from common slopes.

$$\implies E[Y_{1i} | X_i] = \tau + E[Y_{0i} | X_i] = \tau + \alpha + \beta X_i$$

Recall our definition of  $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$ .

$$\begin{aligned} E[Y_i | X_i, D_i] &= D_i E[Y_{1i} | X_i] + (1 - D_i) E[Y_{0i} | X_i] \\ &= \alpha + \tau D_i + \beta X_i = \alpha + \tau D_i + \beta (\tilde{X}_i + c) = \tilde{\alpha} + \tau D_i + \beta \tilde{X}_i \end{aligned}$$

which we can estimate by regressing  $Y_i$  on  $D_i$  and  $\tilde{X}_i$ .

# Sharp RDs

## Estimation: Linear, differing slopes

Assumption  $E[Y_{0i}|X_i = x]$  and  $E[Y_{1i}|X_i = x]$  are linear in  $x$ , i.e.,  
 $E[Y_{0i} | X_i] = \alpha_0 + \beta_0 X_i$  and  $E[Y_{1i} | X_i] = \alpha_1 + \beta_1 X_i$

Now treatment effects can vary with  $X_i$ .

$$\implies E[Y_{1i} - Y_{0i} | X_i] = (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) X_i$$

$$\begin{aligned} E[Y_i | X_i, D_i] &= D_i E[Y_{1i} | X_i] + (1 - D_i) E[Y_{0i} | X_i] \\ &= \alpha_0 + \beta_0 X_i + (\alpha_1 - \alpha_0) D_i + (\beta_1 - \beta_0) D_i X_i \\ &= \tilde{\alpha} + \beta_0 \tilde{X}_i + \tau D_i + \tilde{\beta} D_i \tilde{X}_i \end{aligned}$$

$\tau$  is the LATE at  $\tilde{X}_i = 0$  ( $X_i = c$ ). Estimate: Regress  $Y_i$  in  $\tilde{X}_i$ ,  $D_i$ , and  $D_i \tilde{X}_i$ .<sup>†</sup>

<sup>†</sup> See Appendix for omitted steps.

Fuzzy RDs

# Fuzzy RDs

## Setup

As with their sharp-RD relatives, **fuzzy RDs** take advantage of a discontinuous change in treatment assignment across some threshold  $c$ .

In a **fuzzy regression discontinuity**, the *probability* of treatment changes discontinuously as  $\mathbf{X}_i$  crosses  $c$ , but it is no longer deterministic.

Formally,

$$0 < \lim_{x \downarrow c} \Pr(D_i = 1 \mid \mathbf{X}_i = x) - \lim_{x \uparrow c} \Pr(D_i = 1 \mid \mathbf{X}_i = x) < 1$$

*Ex.*, Exceeding a minimum GRE requirement for graduate school.

# Fuzzy RDs

## Threshold effects

We now have **two effects** of  $X_i$  crossing our threshold  $c$ .

1. The effect of  $X_i$  crossing  $c$  on our outcome

$$\lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x]$$

2. The effect of  $X_i$  crossing  $c$  on the probability of treatment

$$\lim_{x \downarrow c} E[D_i \mid X_i = x] - \lim_{x \uparrow c} E[D_i \mid X_i = x]$$

The treatment effect defined by a fuzzy RD is the ratio of (1) to (2)

$$\tau_{\text{FRD}} = \frac{\lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x]}{\lim_{x \downarrow c} E[D_i \mid X_i = x] - \lim_{x \uparrow c} E[D_i \mid X_i = x]}$$

# Fuzzy RDs

## An old friend

This definition of the fuzzy-RD treatment effect

$$\tau_{\text{FRD}} = \frac{\lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x]}{\lim_{x \downarrow c} E[D_i \mid X_i = x] - \lim_{x \uparrow c} E[D_i \mid X_i = x]}$$

should remind you of something—**IV**, where  $Z_i = \mathbb{I}\{X_i \geq c\}$ .

Accordingly, fuzzy RDs are going to have the **same requirements and interpretation as IV**.

# Fuzzy RDs

## More formally

Let  $D_i(x^*)$  denote the **potential treatment status** of  $i$  **with threshold**  $x^*$ .

Why write potential treatment status  $D_i$  a function of the threshold?

Changing the threshold (e.g., voting age) generally makes more sense than changing  $X_i$  (e.g., age).<sup>†</sup>

*I.e.*, changing the threshold changes treatment statuses at the marginal.

*Assumption*  $D_i(x^*)$  is non-increasing in  $x^*$  at  $x^* = c$ .

This is our monotonicity assumption for fuzzy RDs. If we raise  $x^*$  from  $c$  to  $c + \epsilon$ , no one joins treatment—no defiers.

<sup>†</sup> This observation/motivation can help with inference.

# Fuzzy RDs

## Compliance

Our **compliers** in this setting are individuals such that

$$\lim_{x^* \downarrow X_i} D_i(x^*) = 0 \quad \lim_{x^* \uparrow X_i} D_i(x^*) = 1$$

i.e., **compliers** are only treated when  $x^*$  (the threshold) is *below* their  $X_i$ .

Back to the fuzzy RD treatment effect

$$\begin{aligned} \tau_{\text{FRD}} &= \frac{\lim_{x \downarrow c} E[Y_i \mid X_i = x] - \lim_{x \uparrow c} E[Y_i \mid X_i = x]}{\lim_{x \downarrow c} E[D_i \mid X_i = x] - \lim_{x \uparrow c} E[D_i \mid X_i = x]} \\ &= E[\mathbf{Y}_{1i} - \mathbf{Y}_{0i} \mid i \text{ is a complier and } X_i = c]. \end{aligned}$$

Thus,  $\tau_{\text{FRD}}$  can be a *very local* LATE.

# Graphical analysis

# Graphical analysis

## General

**RD analyses hinge on their graphical analyses.**

If the discontinuity is not graphically apparent, most people are not going to care about the results of a few tortured regressions.

You're arguing you know that treatment assignment changes across the threshold. If your reader/viewer cannot see it, they're likely not going to believe your regression tables.<sup>†</sup>

<sup>†</sup> This skepticism may be well founded. We know RDs are sensitive to functional form—and researchers have been known to *p-hack*.

# Graphical analysis

## Three figures

Most RD analyses will have some subset of three types of figures.

1. **Outcomes** by the running/forcing variable] ( $X_i$ )  
*Do we observe a treatment effect across the discontinuity?*
2. **Covariates** by the running/forcing variable ( $X_i$ )  
*Are covariates smooth/balanced across the discontinuity?*
3. **Density** of running/forcing variable ( $X_i$ )  
*Is there evidence of sorting into treatment (across the threshold)?*

# Graphical analysis

## Outcomes by running variable

These figures tend to show the average value of the outcome  $\mathbf{Y}_i$  at evenly spaced bins of the running variable  $\mathbf{X}_i$ .

We have two parameter choices

1. Binwidth ( $h$ )
2. Numbers of bins below and above threshold ( $K_0, K_1$ )

that yield  $K = K_0 + K_1$  bins ( $k = 1, \dots, K$ )

$$b_k = c - (K_0 - k + 1) \times h$$

We then calculate summaries for each bin.

# Graphical analysis

## Outcomes by running variable

The bin's **number of observations**,  $N_k$

$$N_k = \sum_{i=1}^N \mathbb{I}\{b_k < \mathbf{X}_i \leq b_{k+1}\}$$

The **average treatment level** in the bin,  $\overline{D}_k$  (for fuzzy RDs)

$$\overline{D}_k = \frac{1}{N_k} \sum_{i=1}^N D_i \times \mathbb{I}\{b_k < \mathbf{X}_i \leq b_{k+1}\}$$

The **average outcome** in the bin,  $\overline{Y}_k$

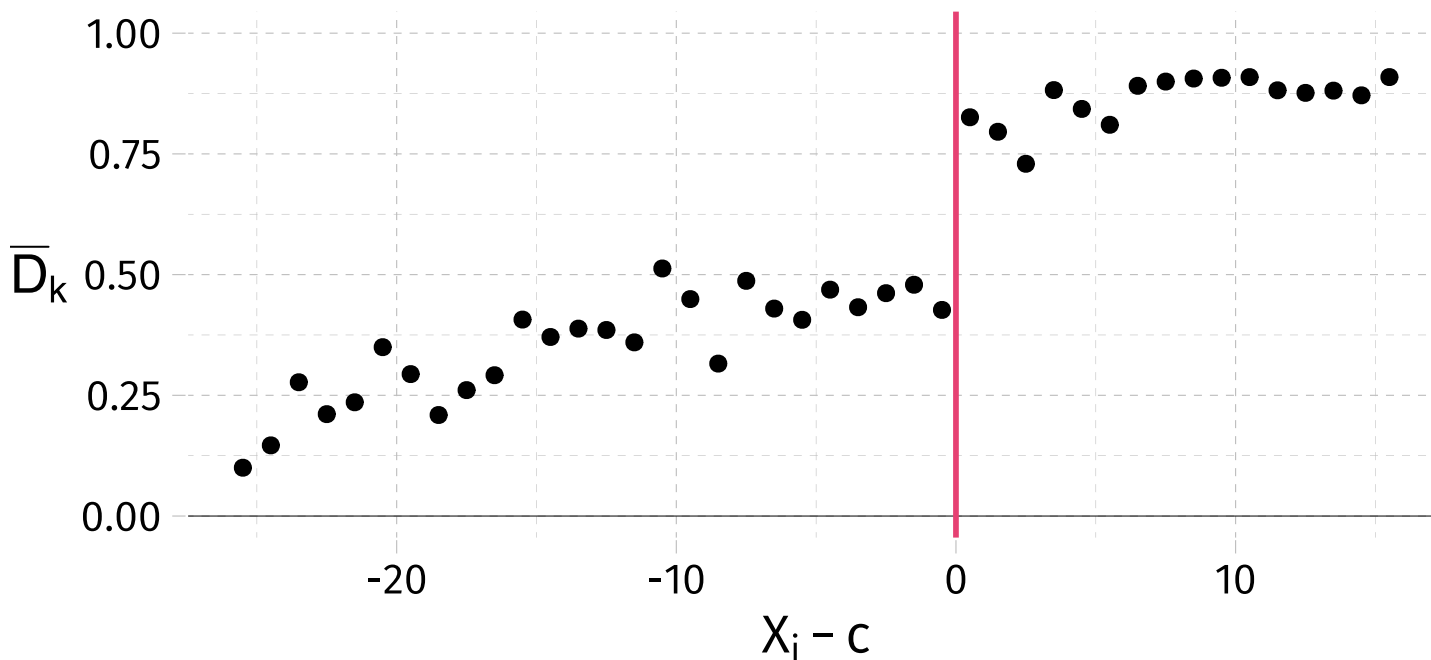
$$\overline{Y}_k = \frac{1}{N_k} \sum_{i=1}^N Y_i \times \mathbb{I}\{b_k < \mathbf{X}_i \leq b_{k+1}\}$$

# Graphical analysis

## Outcomes by running variable

We then plot  $\bar{D}_k$  against the midpoint of each bin.

Q Does crossing  $c$  clearly affect  $\Pr(D_i = 1)$ ? (Fuzzy RD first stage)

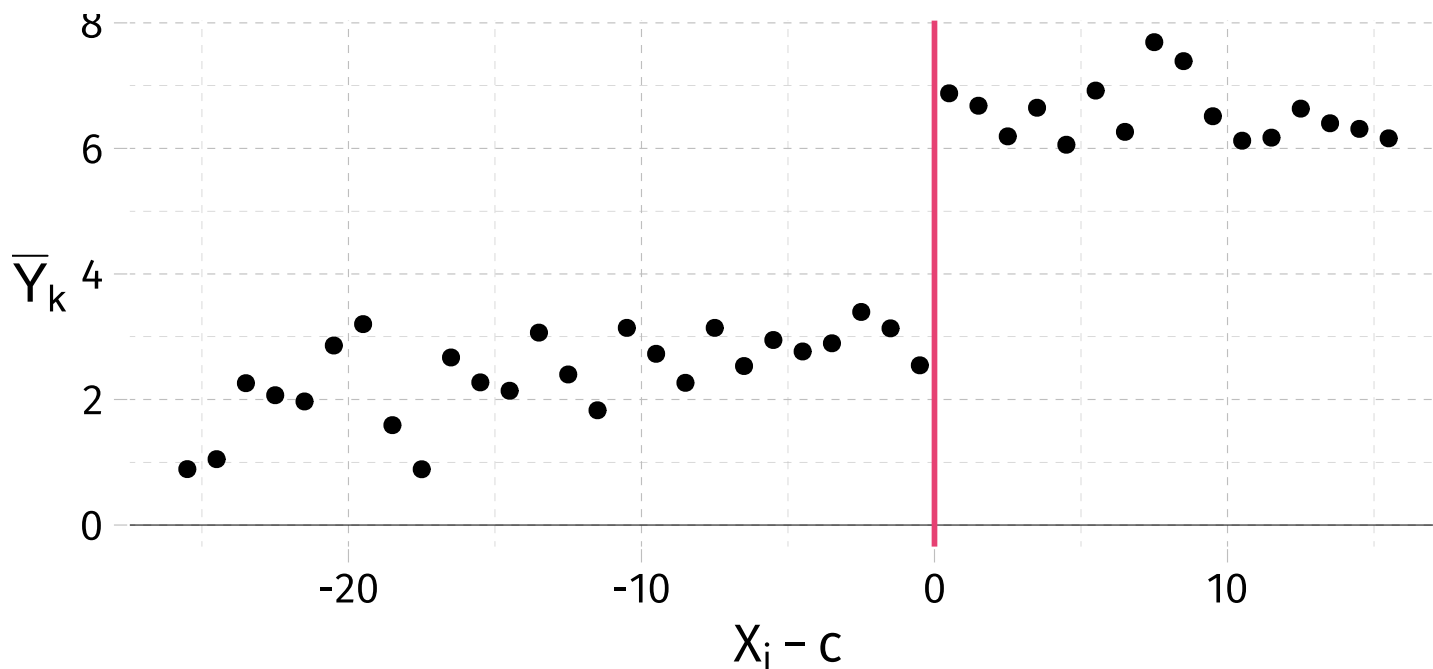


# Graphical analysis

## Outcomes by running variable

And then plot  $\bar{Y}_k$  against the midpoint of each bin.

Q Does crossing  $c$  clearly affect our outcome  $Y_i$ ? (Fuzzy RD reduced form)

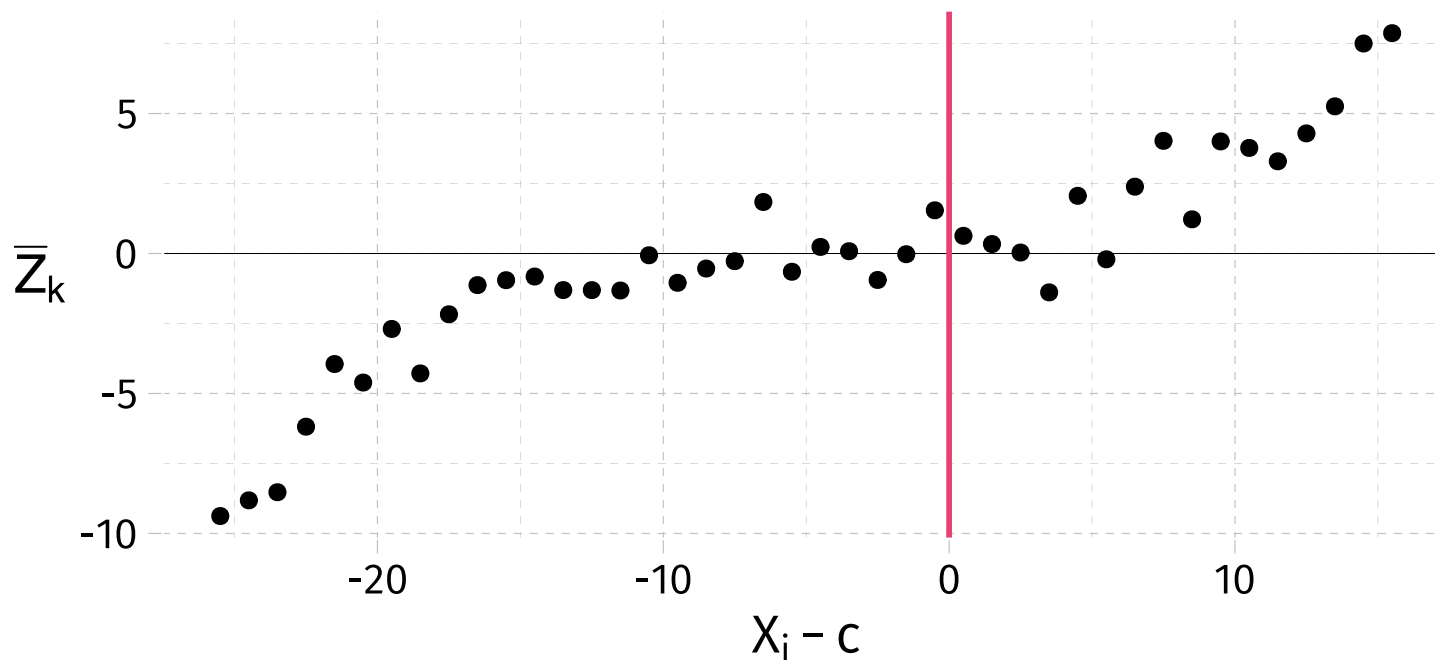


# Graphical analysis

## Covariates by running variable

Now we apply the same approach to covariates ( $Z_i$ ).

Q Are covariates **smooth** across  $c$ ? If not, your RD may be invalid.



# Graphical analysis

## Density of running variable

Finally we looking for other violations of smoothness—particularly in form gaming the threshold.

In other words: Are individuals **bunching** just above or just below the threshold?

If so, folks just below the threshold don't give us the clean counterfactual that we want for the folks just above the threshold.

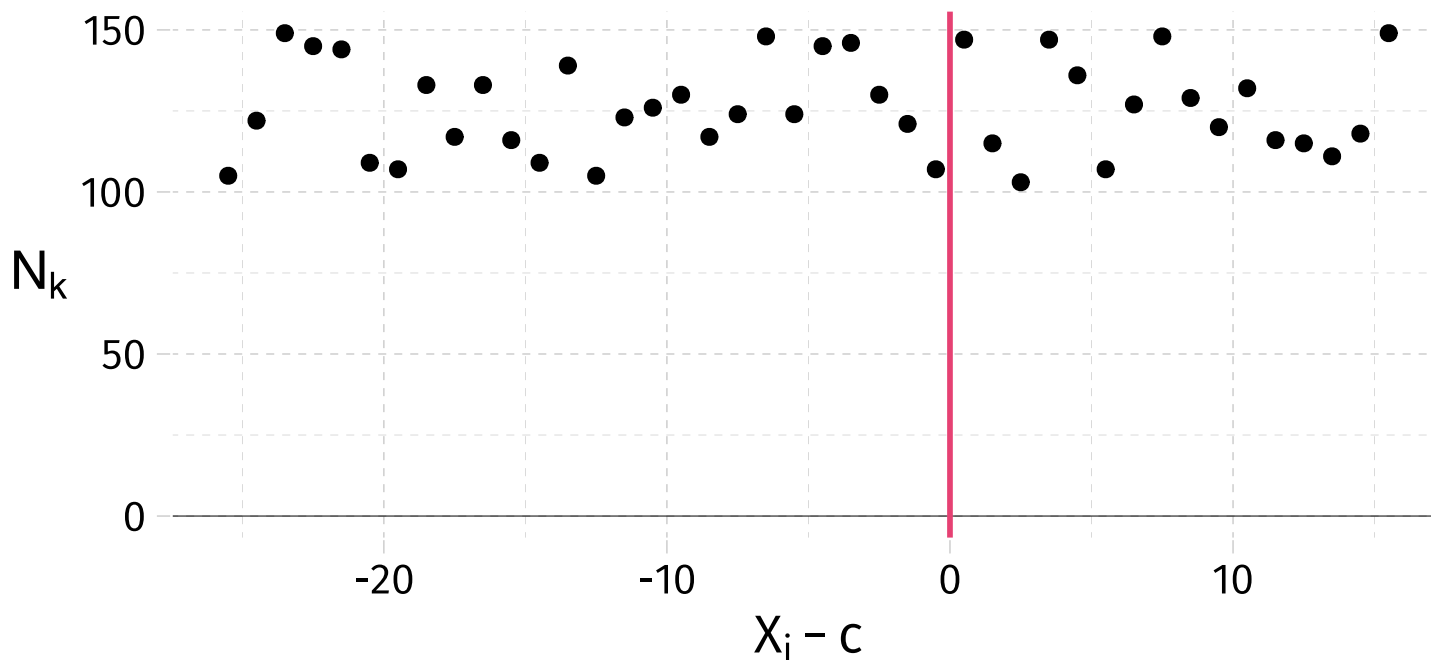
McCrary (2008) suggests testing the density of  $\mathbf{X}_i$  at  $c$ .

# Graphical analysis

## Density of running variable

Effectively, we can plot  $N_k$  at the midpoint of each bin.

Q Is the distribution of  $X_i$  smooth across  $c$ ?

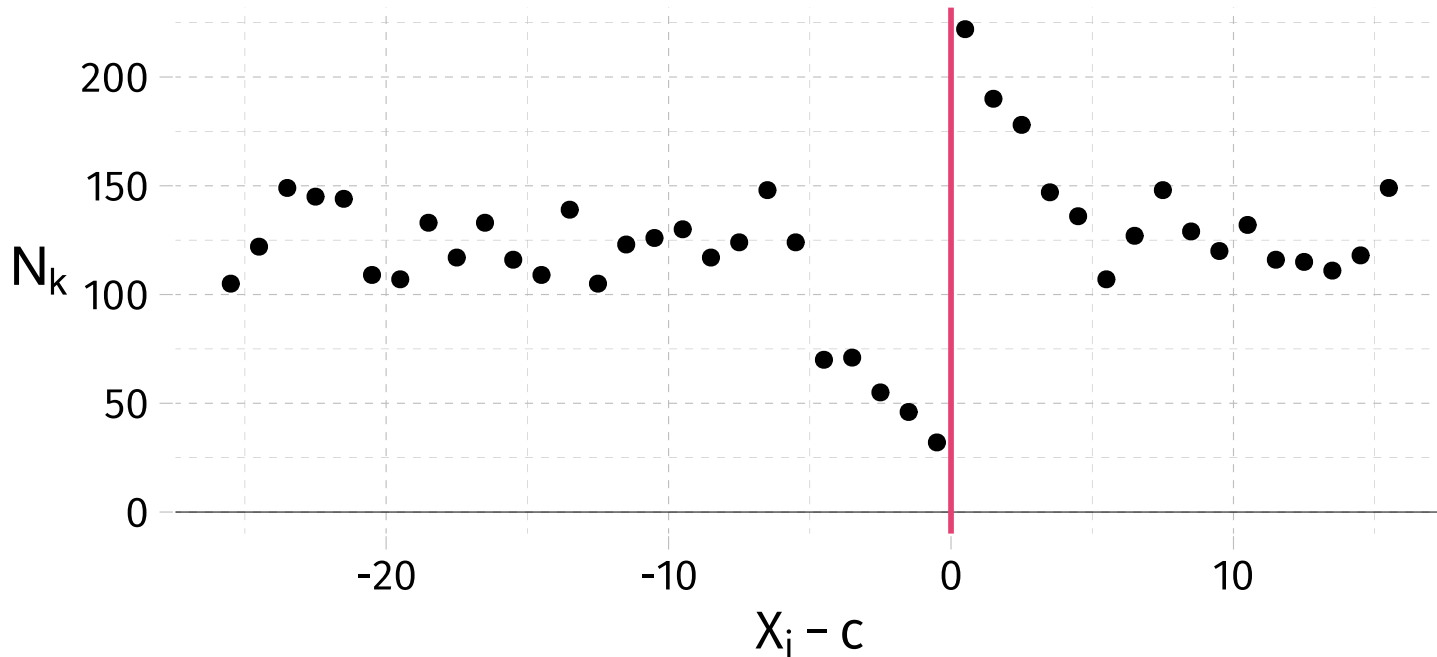


# Graphical analysis

## Density of running variable

**Likely bunching** (problem)

Q Is the distribution of  $X_i$  smooth across  $c$ ?



# Graphical analysis

## Additional points

1. No bin should cross the threshold.
2. Are there discontinuities other than  $c$ ? Should there be? Smoothness?

Again, if these graphs are not clear and convincing, it's going to be hard to make the case that you have a true/credible discontinuity.

# Appendix

## Estimation: Linear, differing slopes

Definitions of terms that magically appear

- $\tilde{\alpha} = \alpha_0 + \beta_0 c$
- $\tau = (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) c$
- $\tilde{\beta} = (\beta_1 - \beta_0)$

# Table of contents

## Admin

1. Schedule

## General RD

1. Setup
2. Framework
3. Examples
4. Sharp vs. fuzzy

## Graphical analysis

1. General
2. Outcomes by  $\mathbf{X}_i$
3. Covariates by  $\mathbf{X}_i$
4. Density of  $\mathbf{X}_i$

## Sharp RDs

1. Setup
2. (Semi) Formally
3. Estimation
4. Examples
5. In practice
6. More estimation

## Fuzzy RDs

1. Setup
2. As IV
3. Somewhat formally