Regression Discontinuity

EC 425/525, Set 9

Edward Rubin

19 May 2019

Prologue

Schedule

Last time

- Introduction to selection-on-unobservables designs
- Instrumental variables
- Two-stage least squares

Today

Regression discontinuity [†]

Upcoming

Midterm

† These notes largely follow notes from Michael Anderson and Imbens and Lemieux (2008).

Regression discontinuity

Regression discontinuity

Setup

We're still in the game of estimating the effect of a potentially endogenous treatment $\mathbf{D}_i$ on an outcome $\mathbf{Y}_i$.

Regression discontinuity

Setup

We're still in the game of estimating the effect of a potentially endogenous treatment $\mathbf{D}_i$ on an outcome $\mathbf{Y}_i$.

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

Setup

We're still in the game of estimating the effect of a potentially endogenous treatment $\mathbf{D}_i$ on an outcome $\mathbf{Y}_i$.

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.

Regression discontinuity

Setup

We're still in the game of estimating the effect of a potentially endogenous treatment $\mathbf{D}_i$ on an outcome $\mathbf{Y}_i$.

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.

Our framework

Back to our potential-outcome framework.

Our framework

Back to our potential-outcome framework.

We want to know the effect of D_i on Y_i.

$$\mathrm{Y}_i = \mathrm{D}_i \mathrm{Y}_{1i} + (1 - \mathrm{D}_i)\mathrm{Y}_{0i}$$

Our framework

Back to our potential-outcome framework.

We want to know the effect of D_i on Y_i.

$$\mathrm{Y}_i = \mathrm{D}_i \mathrm{Y}_{1i} + (1 - \mathrm{D}_i)\mathrm{Y}_{0i}$$

New: Suppose D_i is determined[†] by whether some variable X_i crosses a threshold c (the discontinuity).

† At least in part.

Our framework

Back to our potential-outcome framework.

We want to know the effect of $\mathbf{D}_i$ on $\mathbf{Y}_i$.

$$\mathbf{Y}_i = \mathbf{D}_i \mathbf{Y}_{1i} + (1 - \mathbf{D}_i)\mathbf{Y}_{0i}$$

New: Suppose $\mathbf{D}_i$ is determined[†] by whether some variable $\mathbf{X}_i$ crosses a threshold c (the discontinuity).

The variable $\mathbf{X}_i$ need not be randomly assigned—we will assume it is not (*i.e.*, $\mathbf{X}_i$ correlates with $\mathbf{Y}_{0i}$ and $\mathbf{Y}_{1i}$).

† At least in part.

Regression discontinuity

Our framework

Back to our potential-outcome framework.

We want to know the effect of $\mathbf{D}_i$ on $\mathbf{Y}_i$.

$$\mathbf{Y}_i = \mathbf{D}_i \mathbf{Y}_{1i} + (1 - \mathbf{D}_i)\mathbf{Y}_{0i}$$

New: Suppose $\mathbf{D}_i$ is determined[†] by whether some variable $\mathbf{X}_i$ crosses a threshold $\boldsymbol{c}$ (the discontinuity).

The variable $\mathbf{X}_i$ need not be randomly assigned—we will assume it is not (*i.e.*, $\mathbf{X}_i$ correlates with $\mathbf{Y}_{0i}$ and $\mathbf{Y}_{1i}$).

We will assume that $\mathbf{Y}_{0i}$ and $\mathbf{Y}_{1i}$ vary smoothly in $\mathbf{X}_i$.

† At least in part.

Regression discontinuity

Examples

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

Regression discontinuity

Examples

We often apply regression-discontinuity designs in setting with some arbitrary threshold embeded 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 μg/m^3 are *out of attainment*.

Regression discontinuity

Examples

We often apply regression-discontinuity designs in setting with some arbitrary threshold embeded 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 μg/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 of the threshold is.

Regression discontinuity

Sharp *vs.* fuzzy

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

In **sharp RDs**, individuals move from control to treatment when their $\mathbf{X}_i$ passes our threshold c

Sharp *vs.* fuzzy

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

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

Regression discontinuity

Sharp *vs.* fuzzy

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

In sharp RDs, individuals move from control to treatment when their X_i passes our threshold c, *i.e.*, D_i switches from 0 to 1 when 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.

Regression discontinuity

Sharp *vs.* fuzzy

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

In **sharp RDs**, individuals move from control to treatment when their $\mathbf{X}_i$ passes our threshold c, *i.e.*, 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* $\Pr(\mathrm{D}_i = 1)$ discontinuously jumps at the threshold c, but it does not move from 0 to 1.

Regression discontinuity

Sharp *vs.* fuzzy

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

In **sharp RDs**, individuals move from control to treatment when their $\mathbf{X}_i$ passes our threshold c, *i.e.*, 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* $\mathrm{Pr}(\mathrm{D}_i = 1)$ 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.

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 totally depends upon whether $\mathbf{X}_i \geq c$, *i.e.*,

Sharp RDs

Setup

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

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

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

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 totally depends upon whether $\mathbf{X}_i \geq c$, *i.e.*,

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

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.

More formally

We can write the comparison of means at the threshold as

$$\lim_{x \downarrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x] - \lim_{x \uparrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x]$$

More formally

We can write the comparison of means at the threshold as

$$\lim_{x \downarrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x] - \lim_{x \uparrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x]$$

$$= \lim_{x \downarrow c} E[\mathrm{Y}_{1i} \mid \mathrm{X}_i = x] - \lim_{x \uparrow c} E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x]$$

Sharp RDs

More formally

We can write the comparison of means at the threshold as

$$\lim_{x \downarrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x] - \lim_{x \uparrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x]$$

$$= \lim_{x \downarrow c} E[\mathrm{Y}_{1i} \mid \mathrm{X}_i = x] - \lim_{x \uparrow c} E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x]$$

$$= \tau_{\mathrm{SRD}}$$

Sharp RDs

More formally

We can write the comparison of means at the threshold as

$$\lim_{x\downarrow c} E[\text{Y}_i \mid \text{X}_i = x] - \lim_{x\uparrow c} E[\text{Y}_i \mid \text{X}_i = x]$$

$$= \lim_{x\downarrow c} E[\text{Y}_{1i} \mid \text{X}_i = x] - \lim_{x\uparrow c} E[\text{Y}_{0i} \mid \text{X}_i = x]$$

$$= \tau_{\text{SRD}}$$

Assumption $E[\text{Y}_{1i} \mid \text{X}_i = x]$ and $E[\text{Y}_{0i} \mid \text{X}_i = x]$ are continuous in x.

More formally

We can write the comparison of means at the threshold as

$$\lim_{x\downarrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x] - \lim_{x\uparrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x]$$

$$= \lim_{x\downarrow c} E[\mathrm{Y}_{1i} \mid \mathrm{X}_i = x] - \lim_{x\uparrow c} E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x]$$

$$= \tau_{\text{SRD}}$$

Assumption $E[\mathrm{Y}_{1i} \mid \mathrm{X}_i = x]$ and $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x]$ are continuous in x.

$$\implies \tau_{\text{SRD}} = E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{X}_i = c]$$

Sharp RDs

More formally

We can write the comparison of means at the threshold as

$$\lim_{x\downarrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x] - \lim_{x\uparrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x]$$

$$= \lim_{x\downarrow c} E[\mathrm{Y}_{1i} \mid \mathrm{X}_i = x] - \lim_{x\uparrow c} E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x]$$

$$= \tau_{\text{SRD}}$$

Assumption $E[\mathrm{Y}_{1i} \mid \mathrm{X}_i = x]$ and $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x]$ are continuous in x.

$$\implies \tau_{\text{SRD}} = E[\mathrm{Y}_{1i} - \mathrm{Y}_{0i} \mid \mathrm{X}_i = c]$$

I.e., Because we don't observe Y_{0i} for treated individuals, we extrapolate $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = c - \varepsilon]$ to $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i = x + \varepsilon]$ for small ε.

Sharp RDs

Estimation

Thus, we estimate

$$\tau_{\mathrm{SRD}} = \lim_{x \downarrow c} E[\mathrm{Y}_i - \mathrm{X}_i = x] - \lim_{x \uparrow c} E[\mathrm{Y}_i \mid \mathrm{X}_i = x]$$

as the diffrence between two regression functions estimated "near" c.

Sharp RDs

Estimation

Thus, we estimate

$$\tau_{\text{SRD}} = \lim_{x \downarrow c} E[\mathbf{Y}_i - \mathbf{X}_i = x] - \lim_{x \uparrow c} E[\mathbf{Y}_i \mid \mathbf{X}_i = x]$$

as the diffrence between two regression functions estimated "near" c.

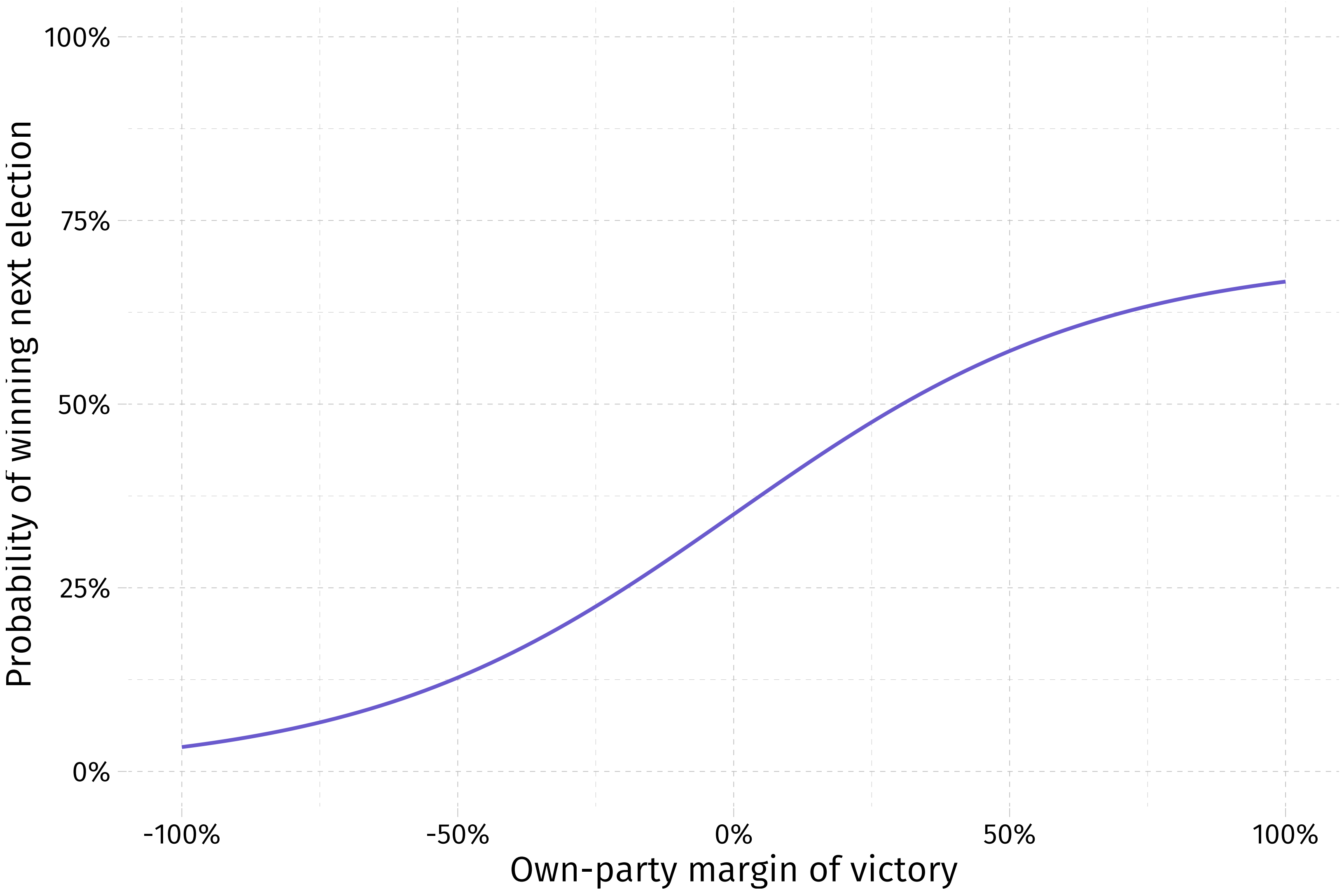
We must stay "near" to c to minimize the bias from extrapolating $E[\mathbf{Y}_{0i} \mid \mathbf{X}_i = c - \varepsilon]$ to $E[\mathbf{Y}_{0i} \mid \mathbf{X}_i = c + \varepsilon]$ (and assuming continuity).

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

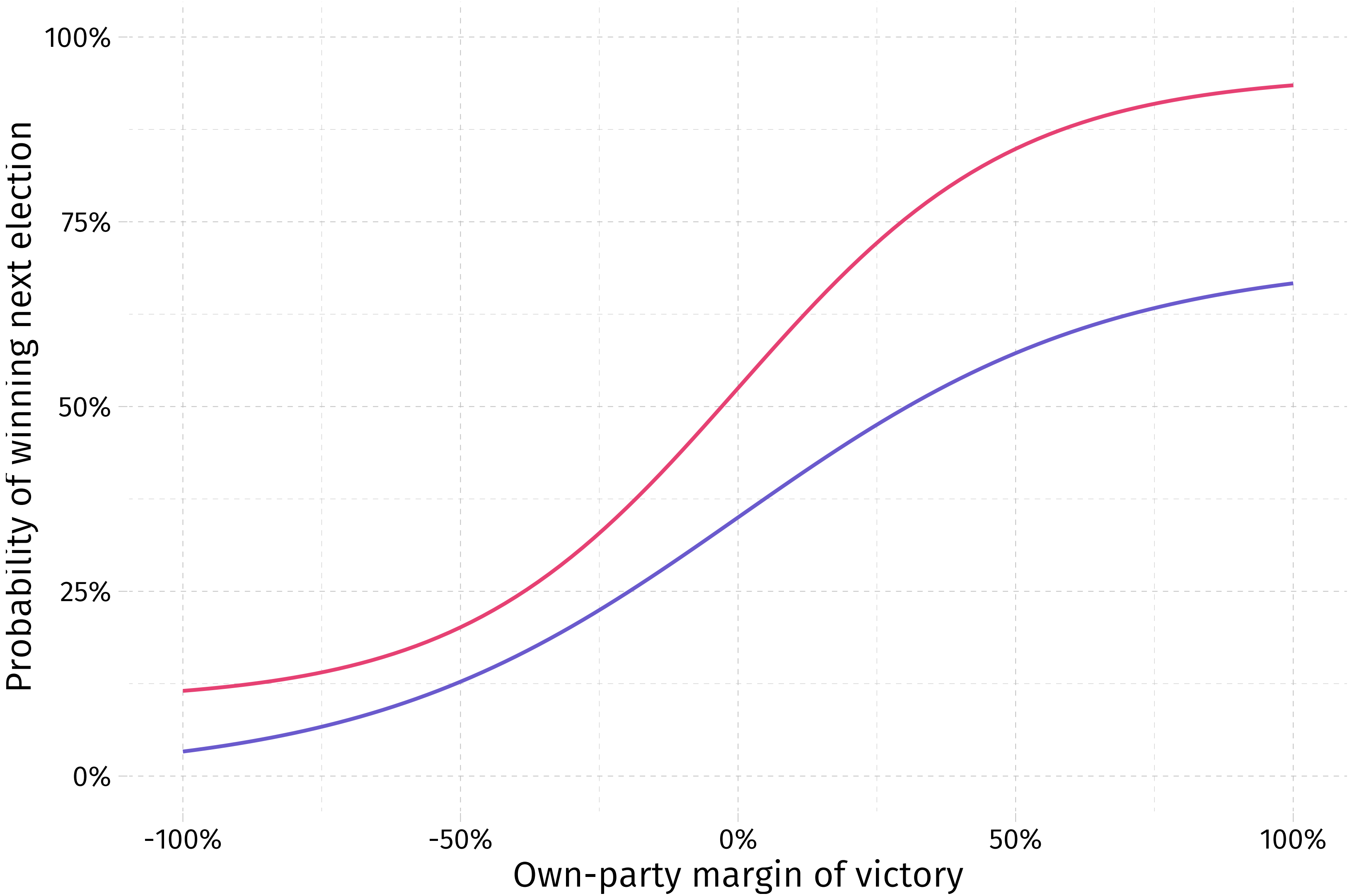
Is there a benefit of incumbency (at the party level)?[†]

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

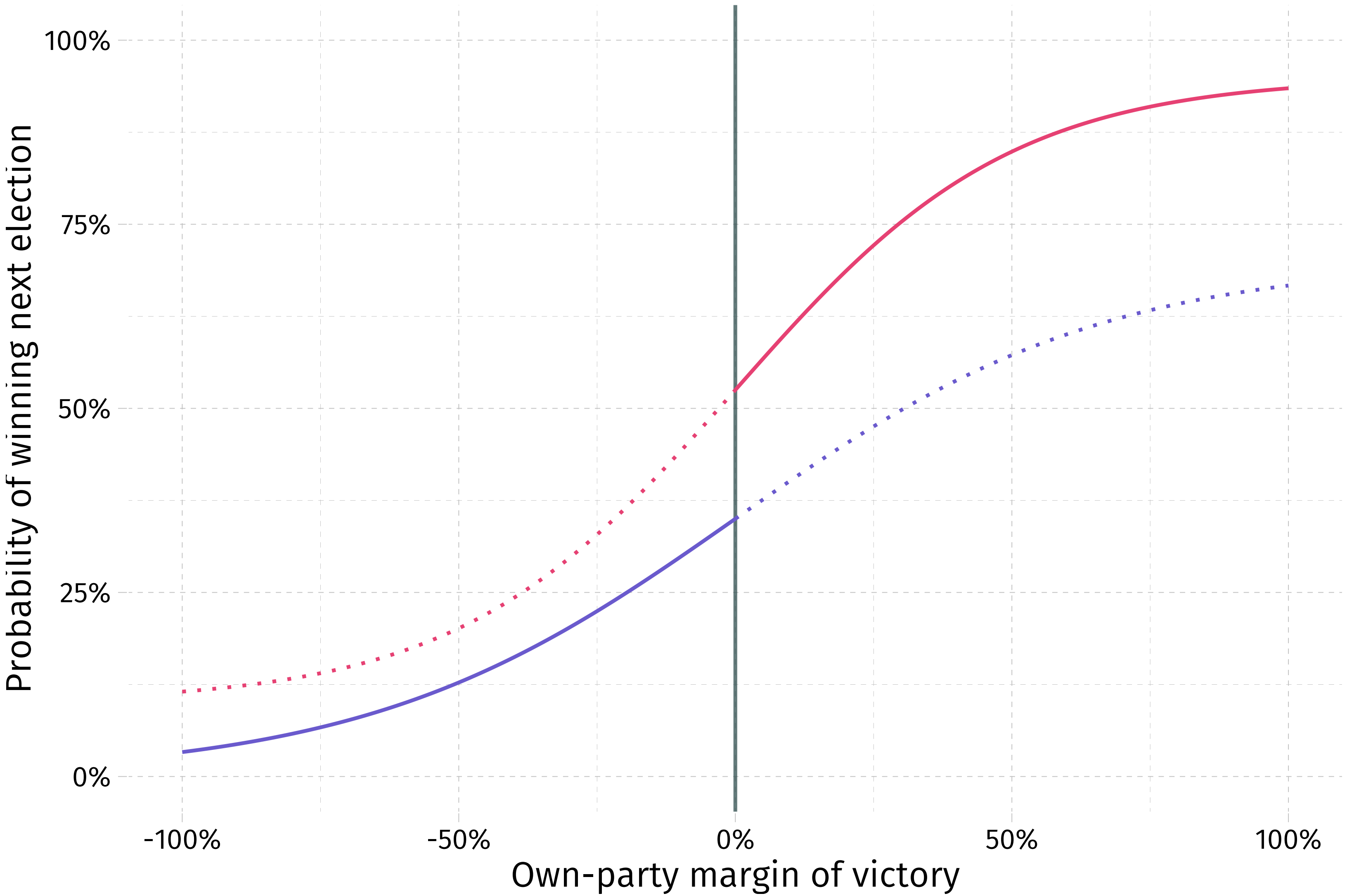
Let's start with $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i]$



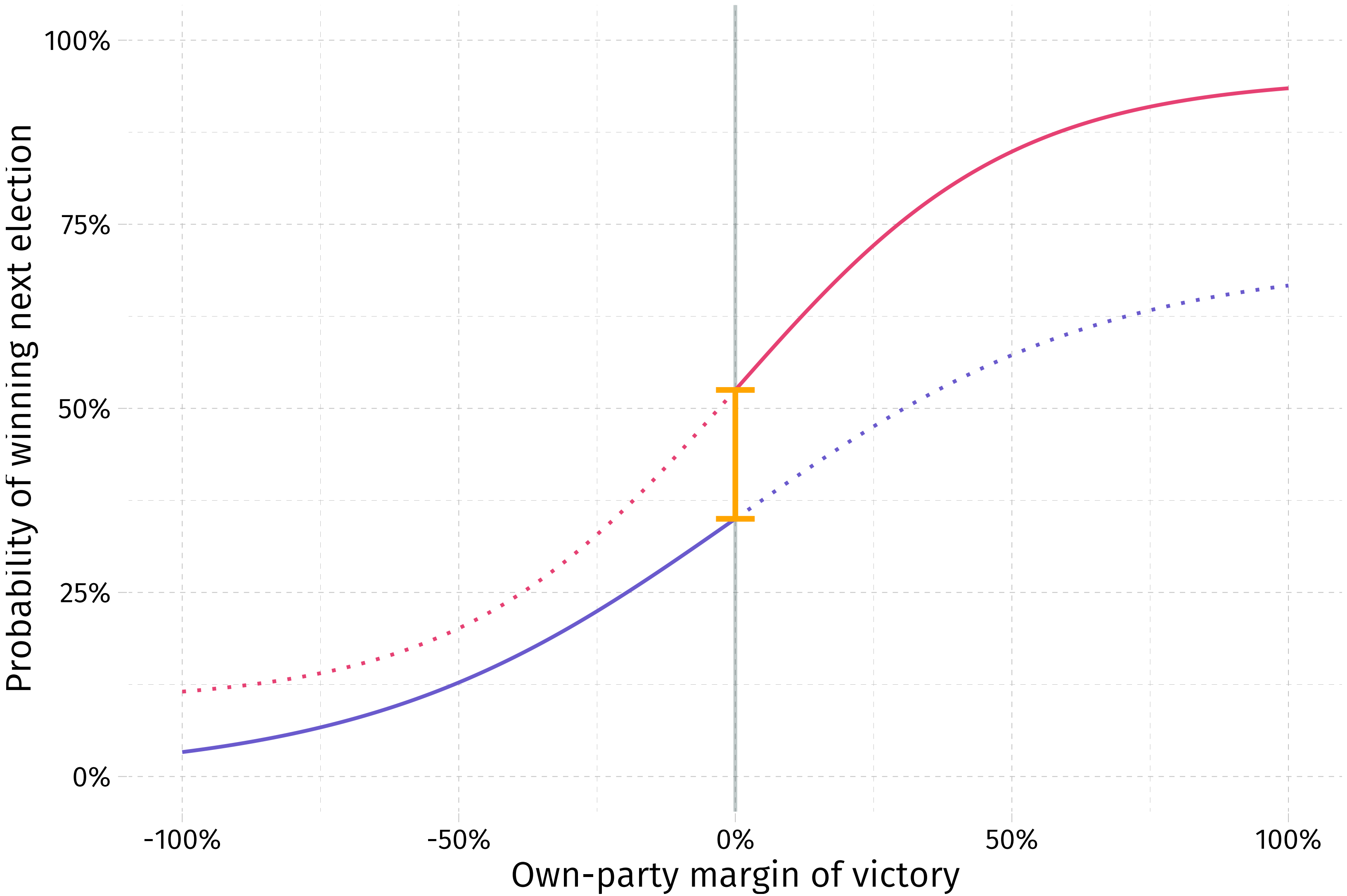
Let's start with $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i]$ and $E[\mathrm{Y}_{1i} \mid \mathrm{X}_i]$.



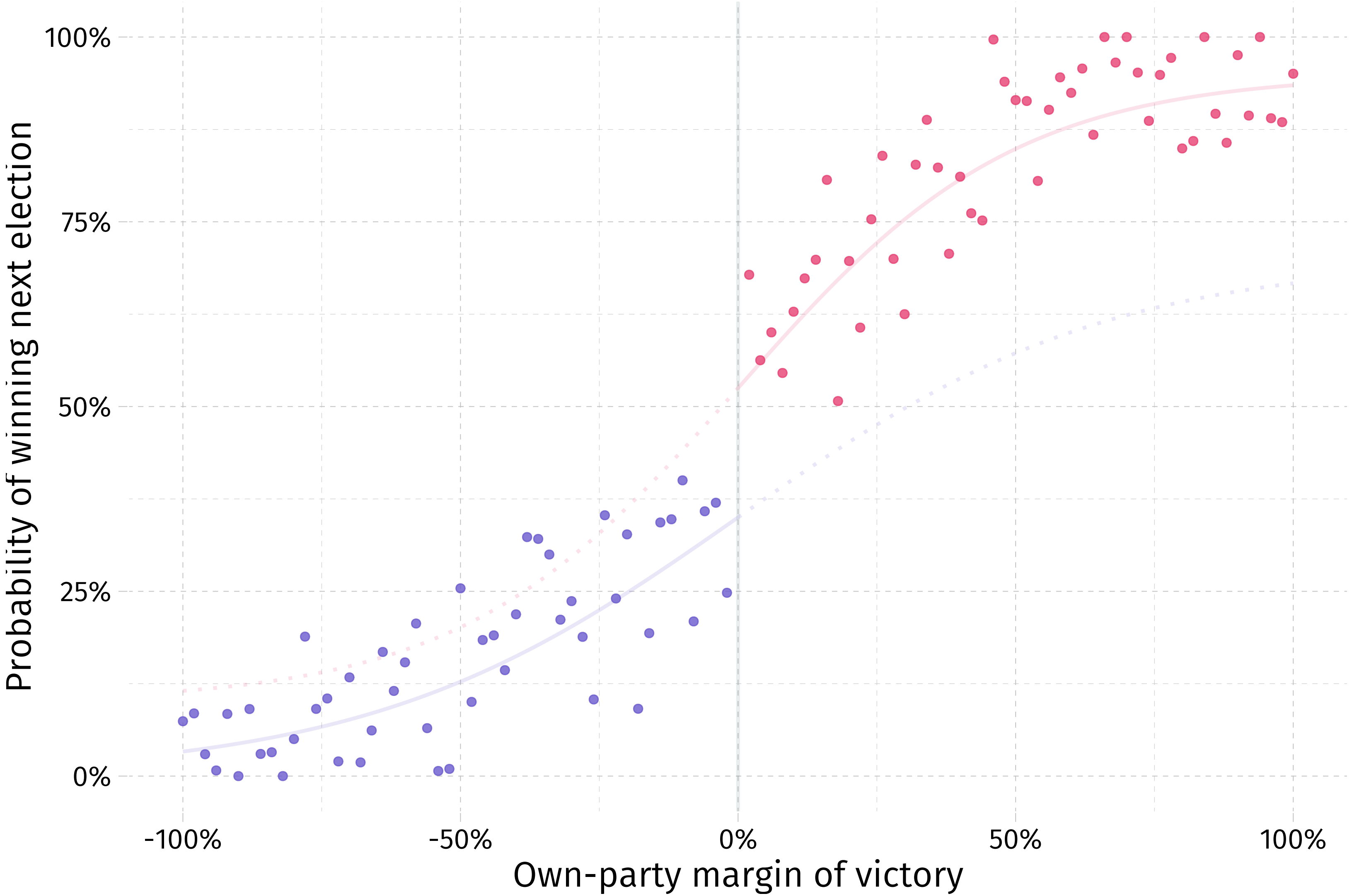
You only win an election if your **margin of victory exceeds zero**.



$E[\mathrm{Y}_{1i} \mid \mathrm{X}_i] - E[\mathrm{Y}_{0i} \mid \mathrm{X}_i]$ **at the discontinuity** gives τ_{SRD}.



Real data are a bit trickier. We must estimate $E[\mathrm{Y}_{1i} \mid \mathrm{X}_i]$ and $E[\mathrm{Y}_{0i} \mid \mathrm{X}_i]$.

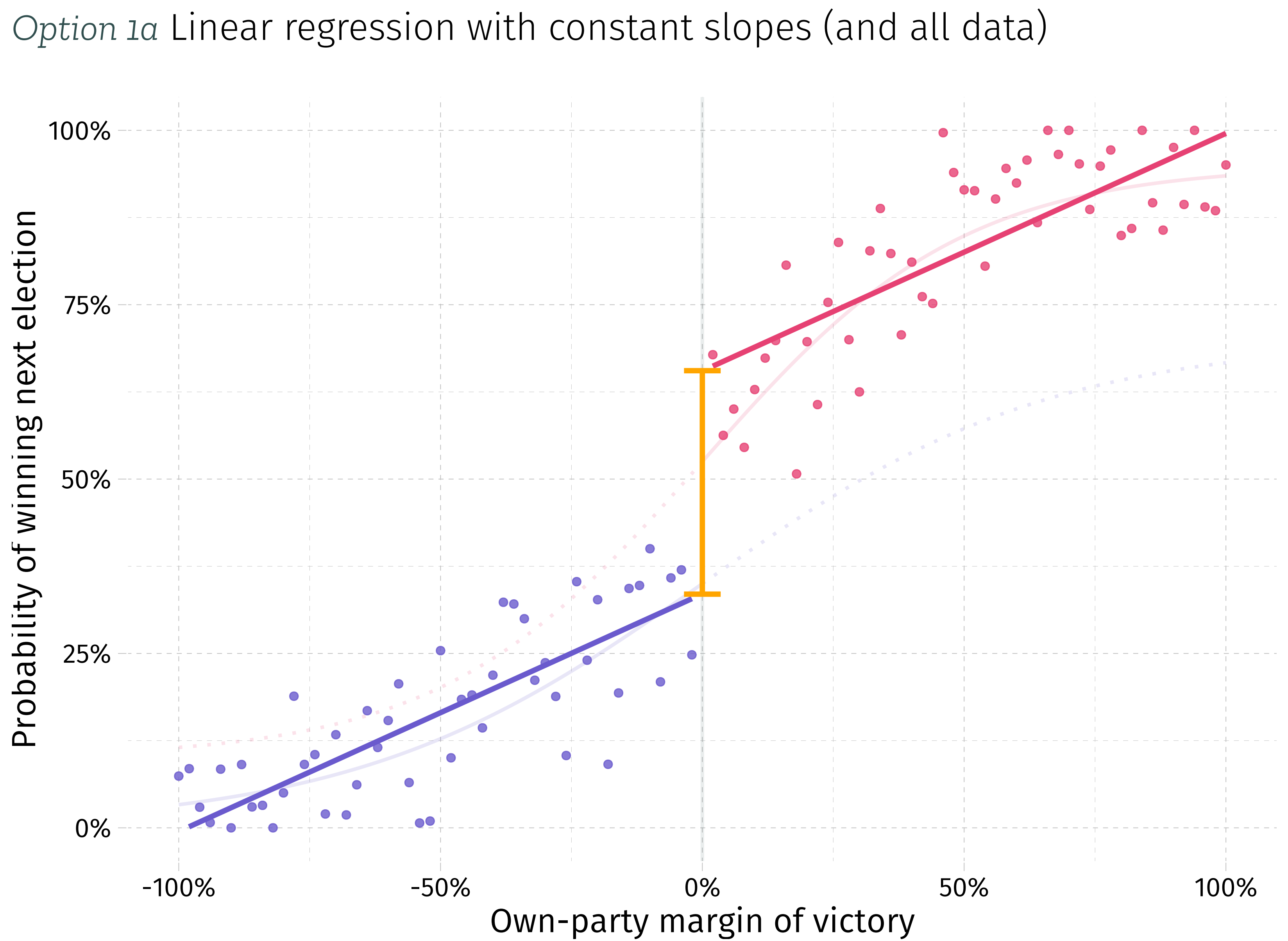


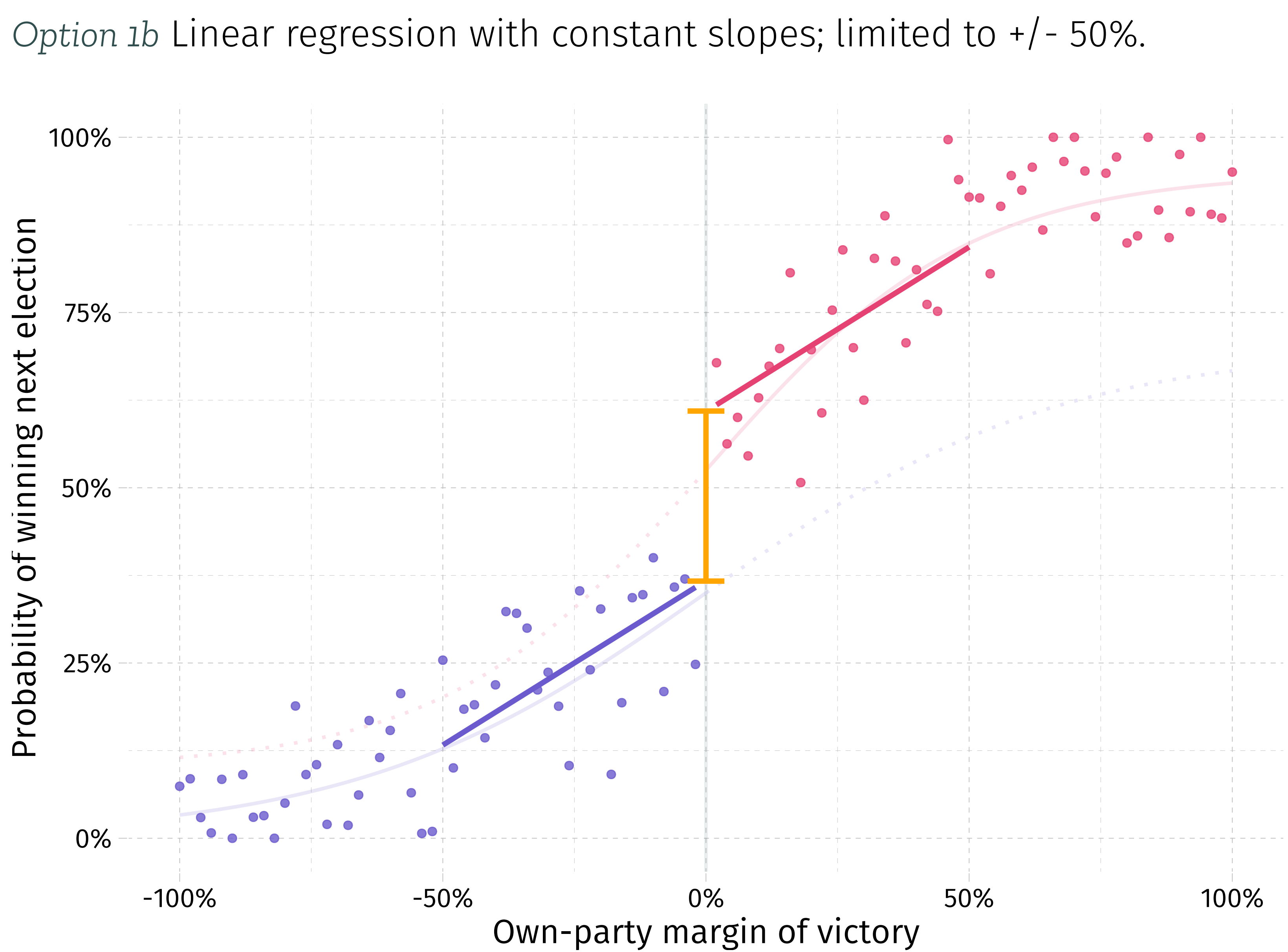
Questions

1. How should we estimate $E[\mathbf{Y}_{1i} \mid \mathbf{X}_i]$ and $E[\mathbf{Y}_{0i} \mid \mathbf{X}_i]$?

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)
Probability of winning next election
100%
75%
50%
25%
0%
-100%
-50%
0%
50%
100%
Own-party margin of victory

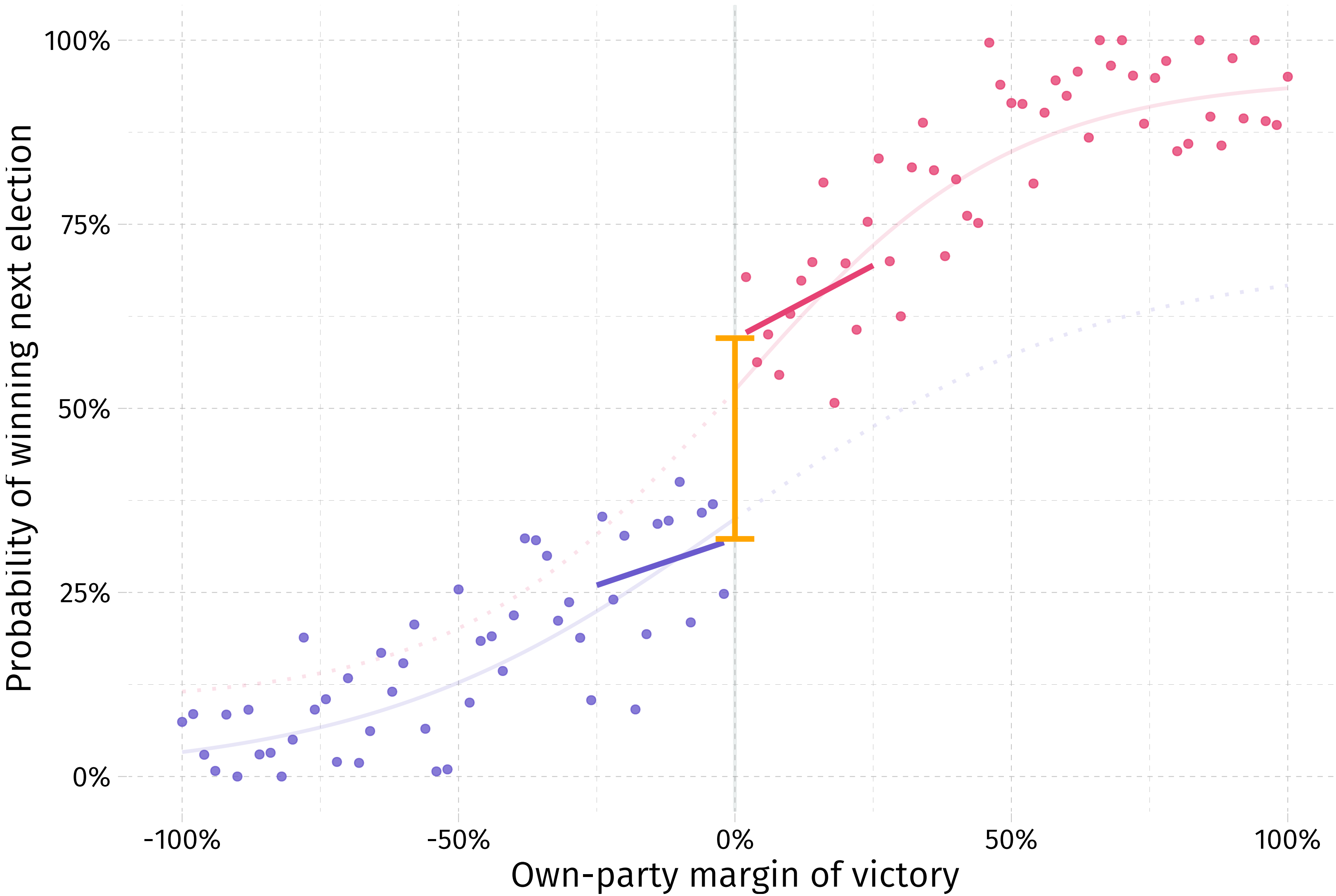


Option 1b Linear regression with constant slopes; limited to +/- 50%.
Probability of winning next election
100%
75%
50%
25%
0%
-100%
-50%
0%
50%
100%
Own-party margin of victory

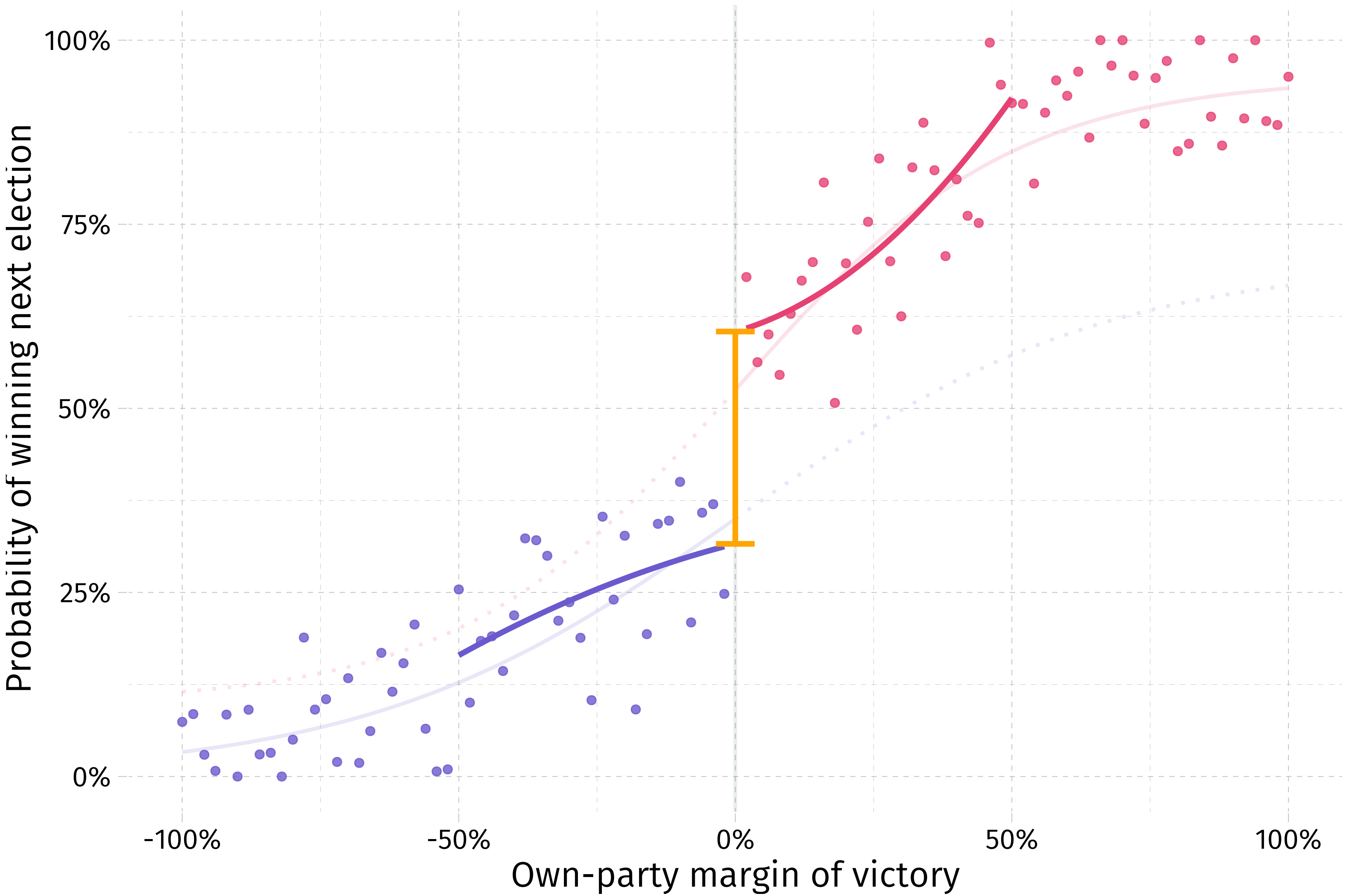
Option 2a Linear regression with differing slopes (and all data)

Option 2b Linear regression with differing slopes; limited to +/- 50%.

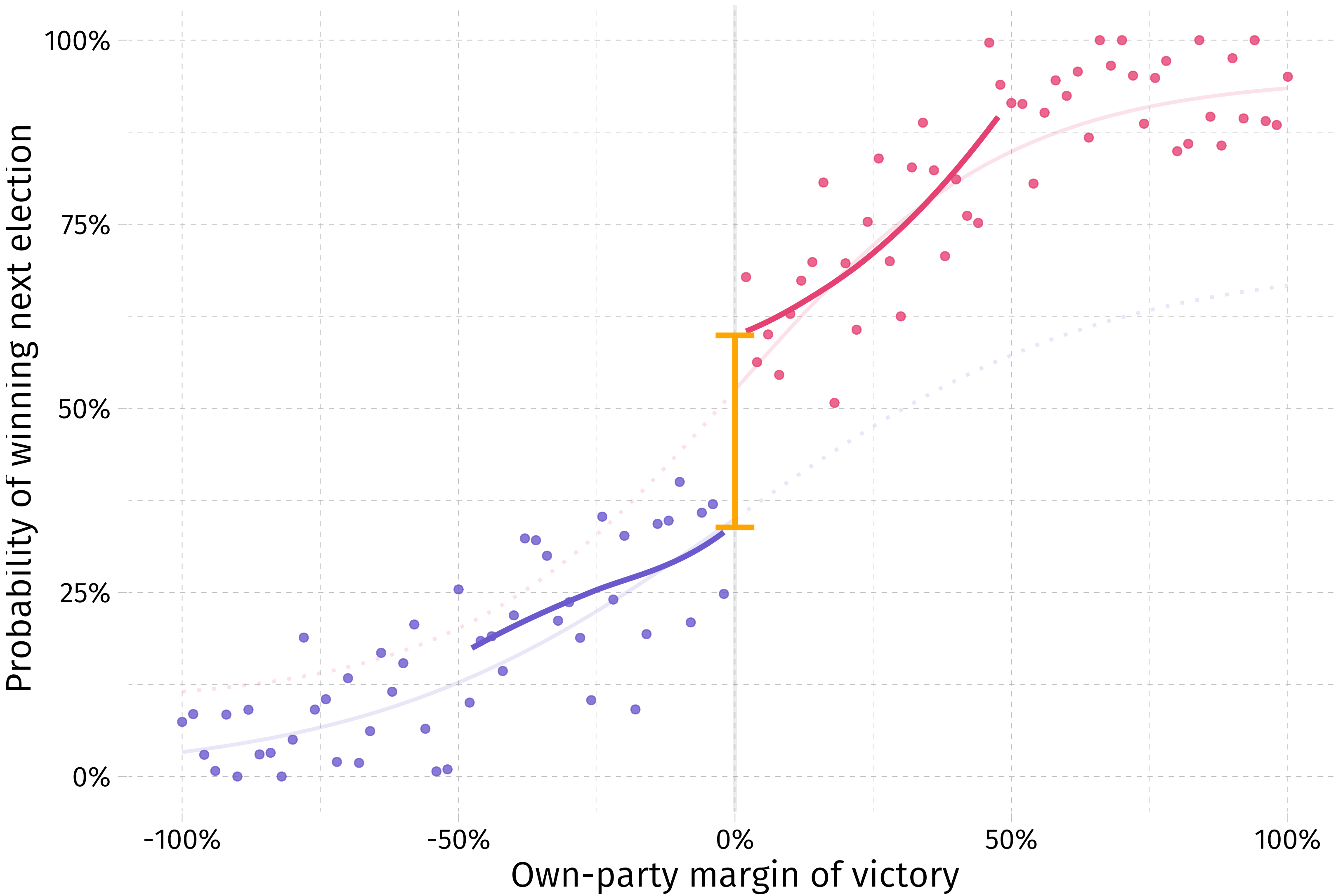
Option 2c Linear regression with differing slopes; limited to +/- 25%.



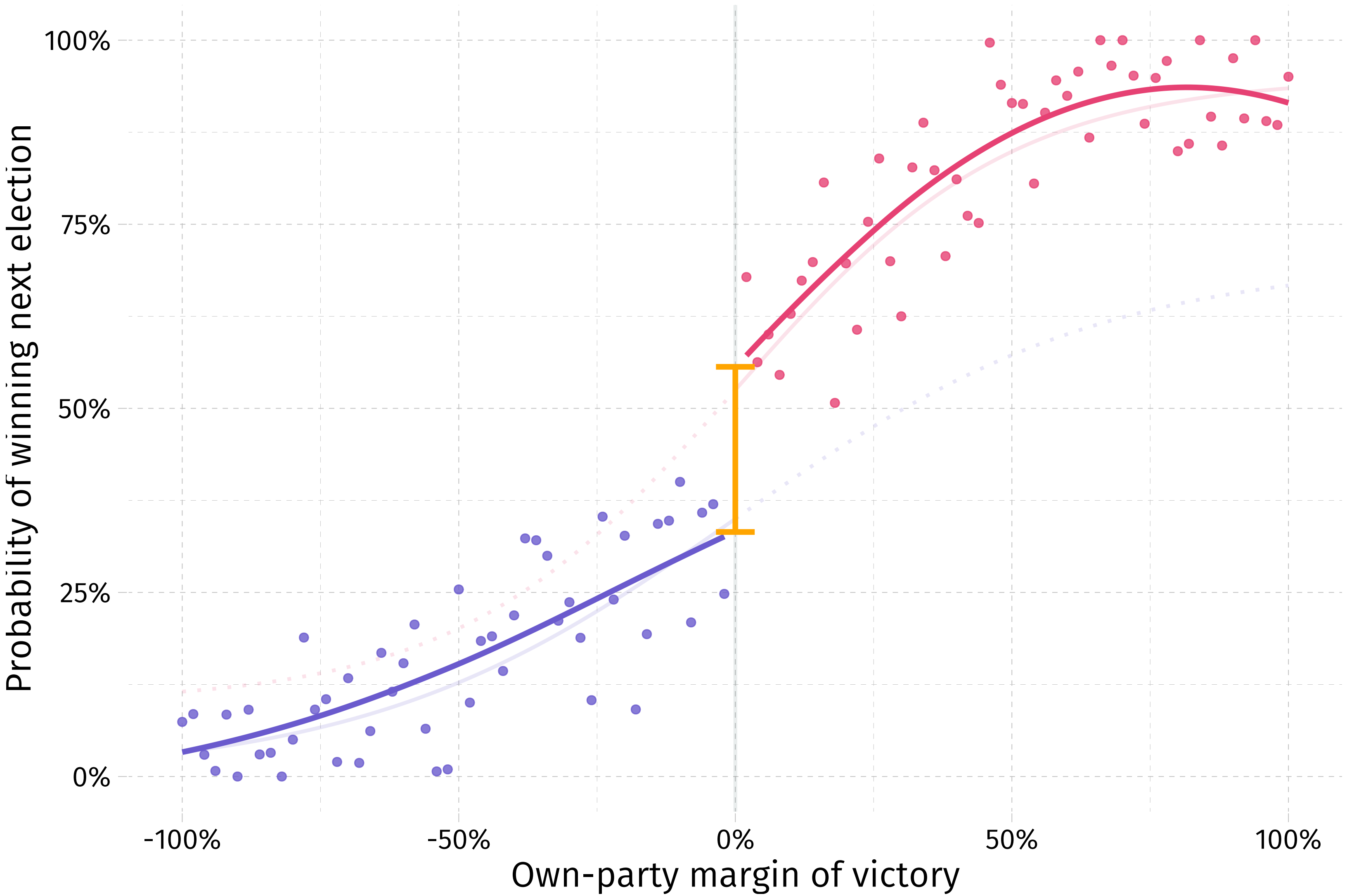
Option 3 Differing quadratic regressions (limited to +/- 50%).



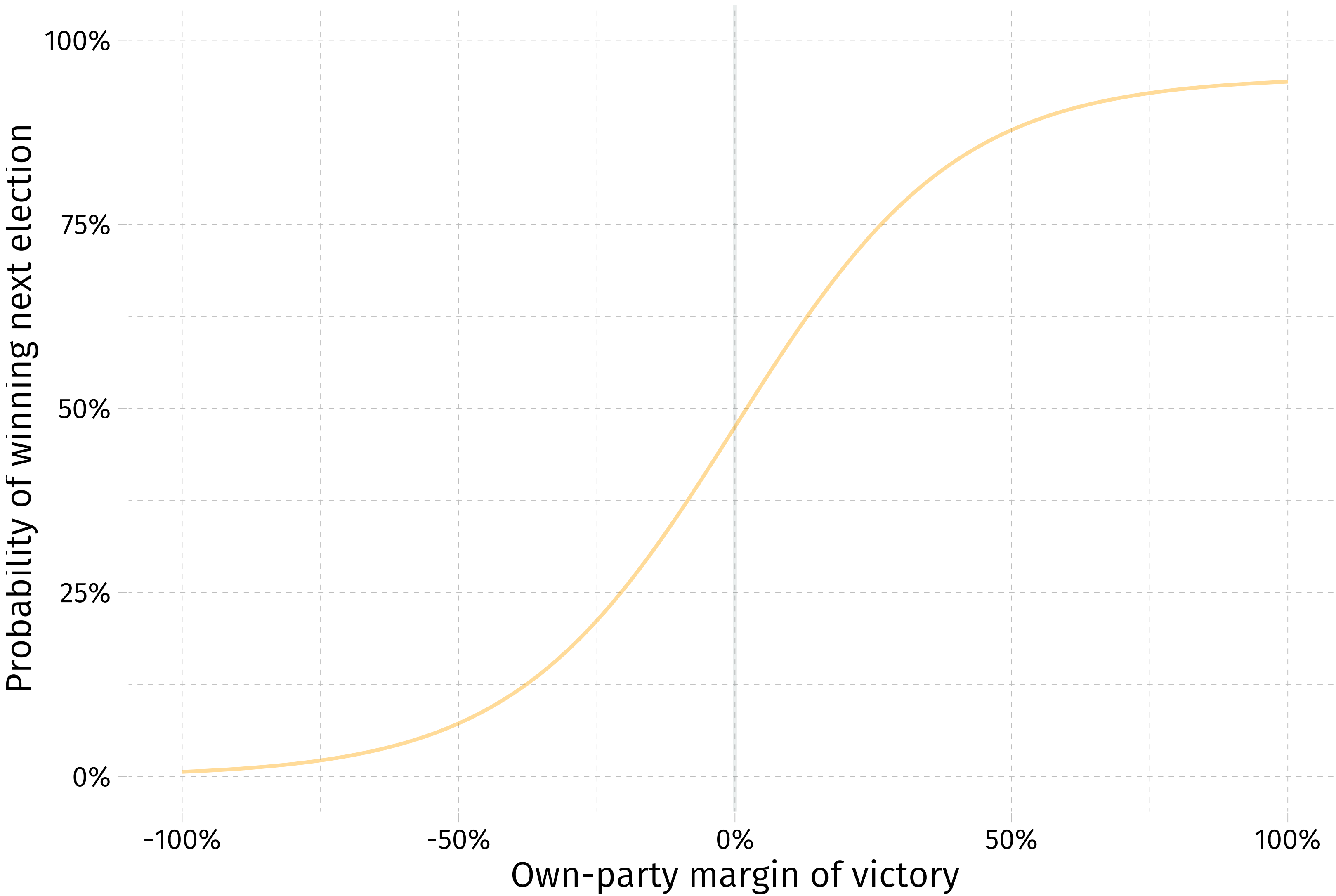
Option 4a Differing local (LOESS) regressions (limited to +/- 50%).



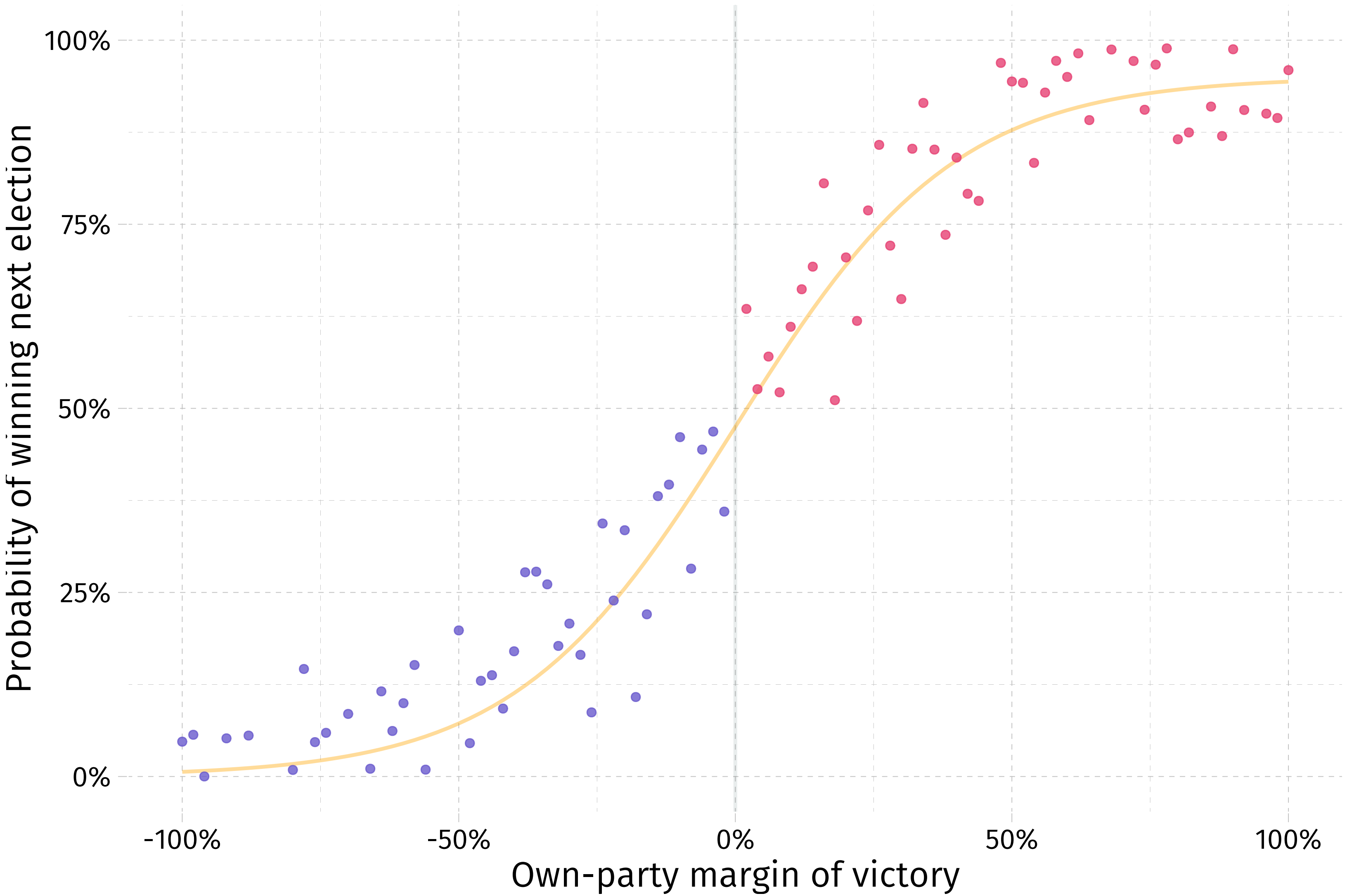
Option 4b Differing local (LOESS) regressions (all data).



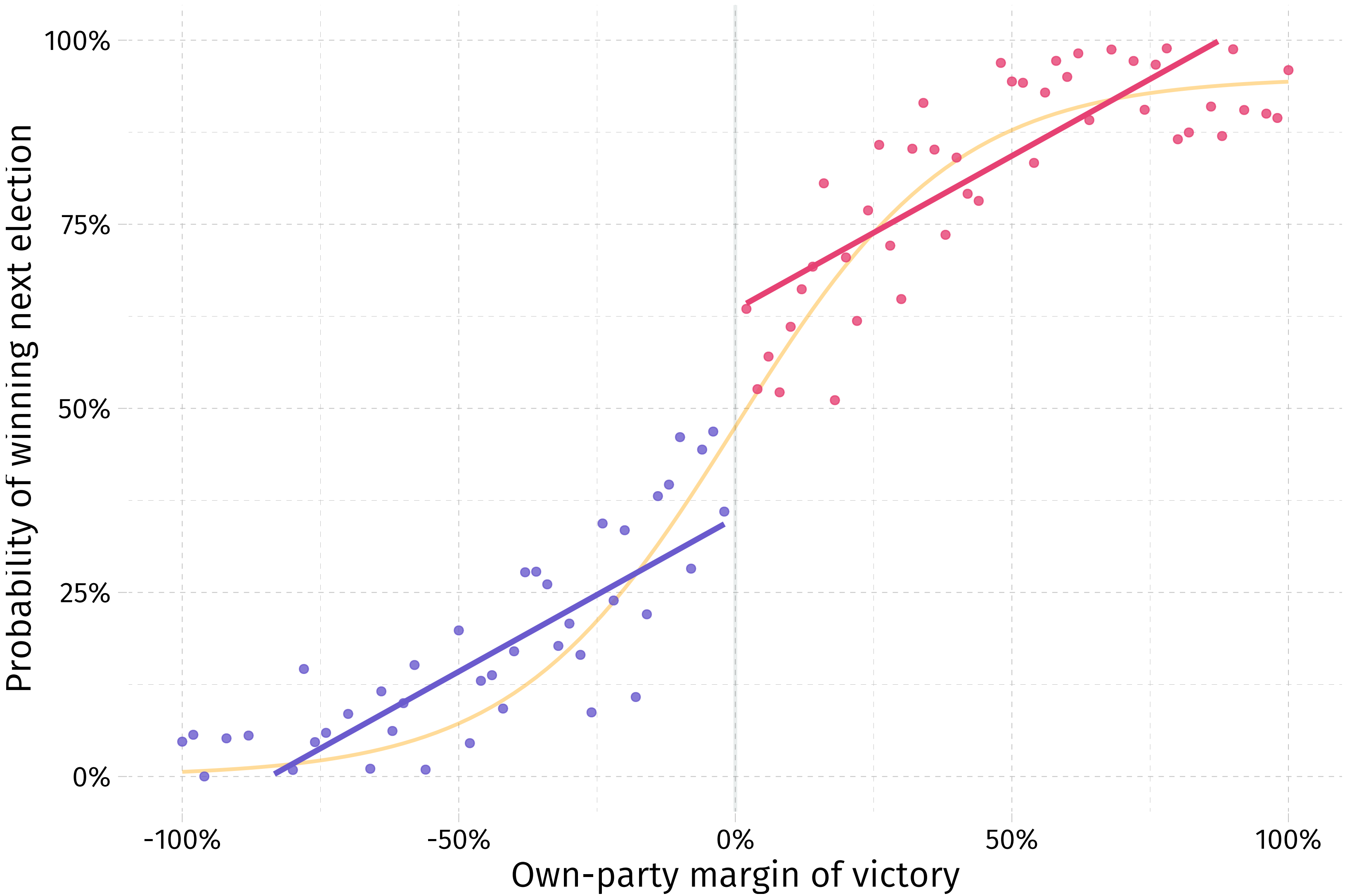
Note Functional form can be very important.



Note Functional form can be very important.

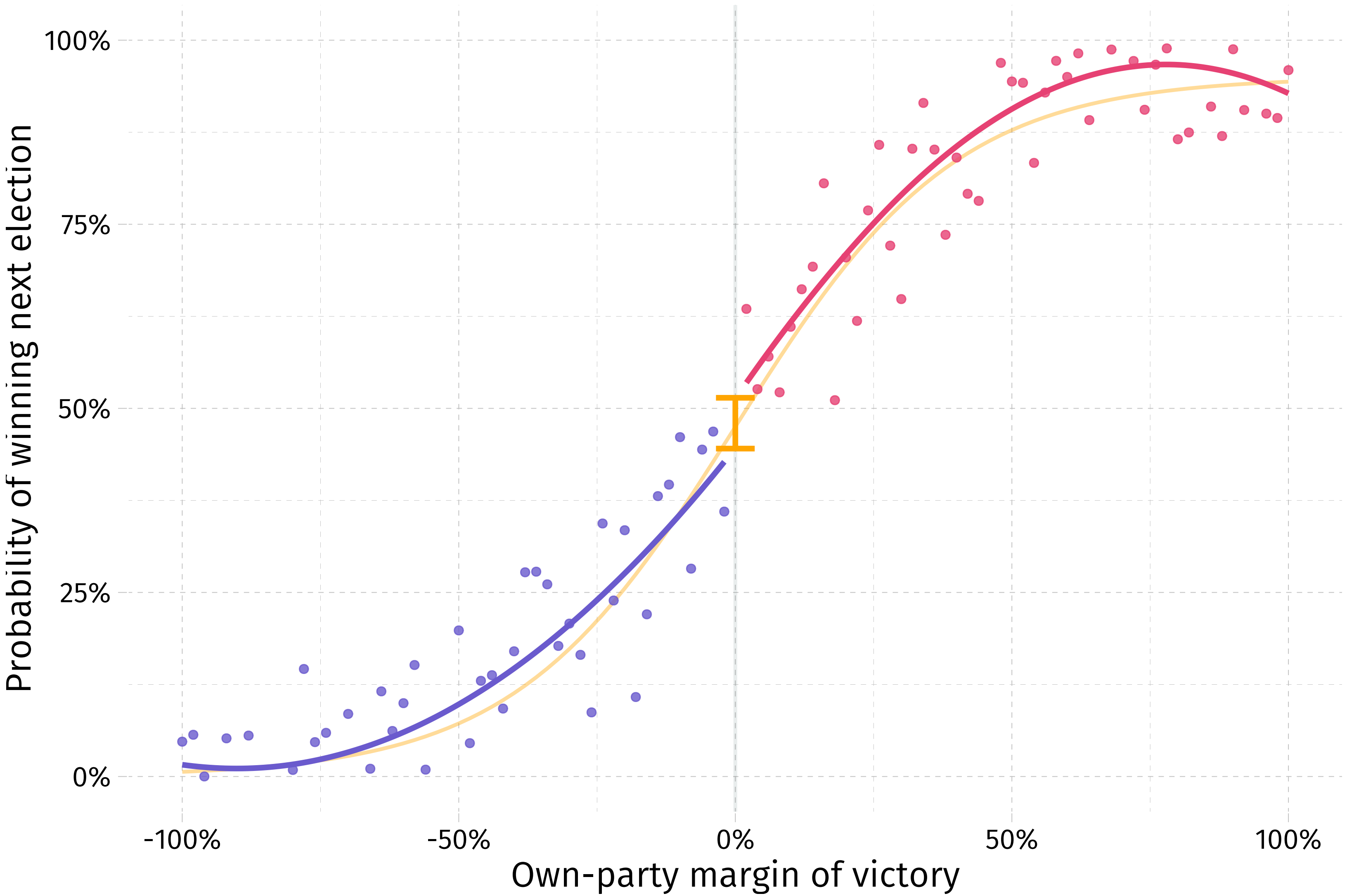


Note Functional form can be very important.

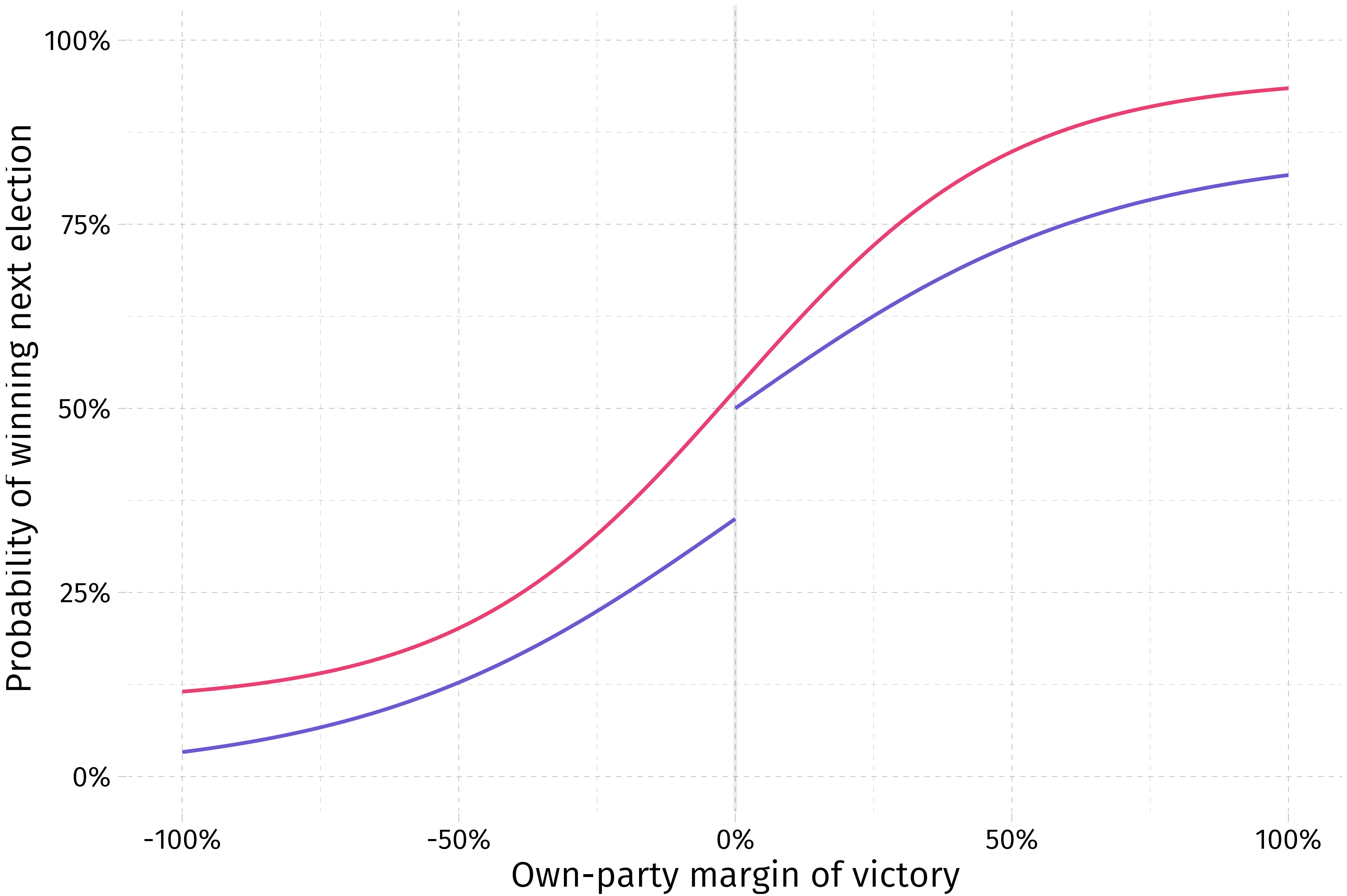


Note Functional form can be very important.

Note Functional form can be very important.



The continuity of $E[\text{Y}_{0i} \mid \text{X}_i = x]$ (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-orderpolynomials 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 othersmooth functions.

Sharp RDs

In practice

Gelman and Imbens (2018) on functional form:

> We argue that controlling for global high-orderpolynomials 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 othersmooth functions.

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

Fuzzy RDs

Fuzzy RDs

Setup

Table of contents

Admin

RD