Optimal Model Selection in RDD and Related Settings Using Placebo Zones

Nathan Kettlewell and Peter Siminski (UTS)

ASSA, January 8, 2023

Graphics created by
Ullupta conecum re volupta
evelignis pe et doluptuam.

Our 2 Stata Programs

ssc install pzms

https://econpapers.repec.org/software/bocbocode/s459073.htm

-pzms- implements our approach. Very easy to use. Only required option is the maximum bandwidth

pzms_sim uses simulations based on the data from any application, to examine likely performance of our approach, compared to alternative approaches https://sites.google.com/site/nrkettlewell/research

Outline

- **Key Issues and our Contribution**
- Motivating application – 2 policy changes affecting Learner drivers in NSW
- Theory – show our approach is asymptotically optimal, under restrictive conditions
- Simulations – our approach performs favourably compared to other procedures using stylised and realistic DGPs

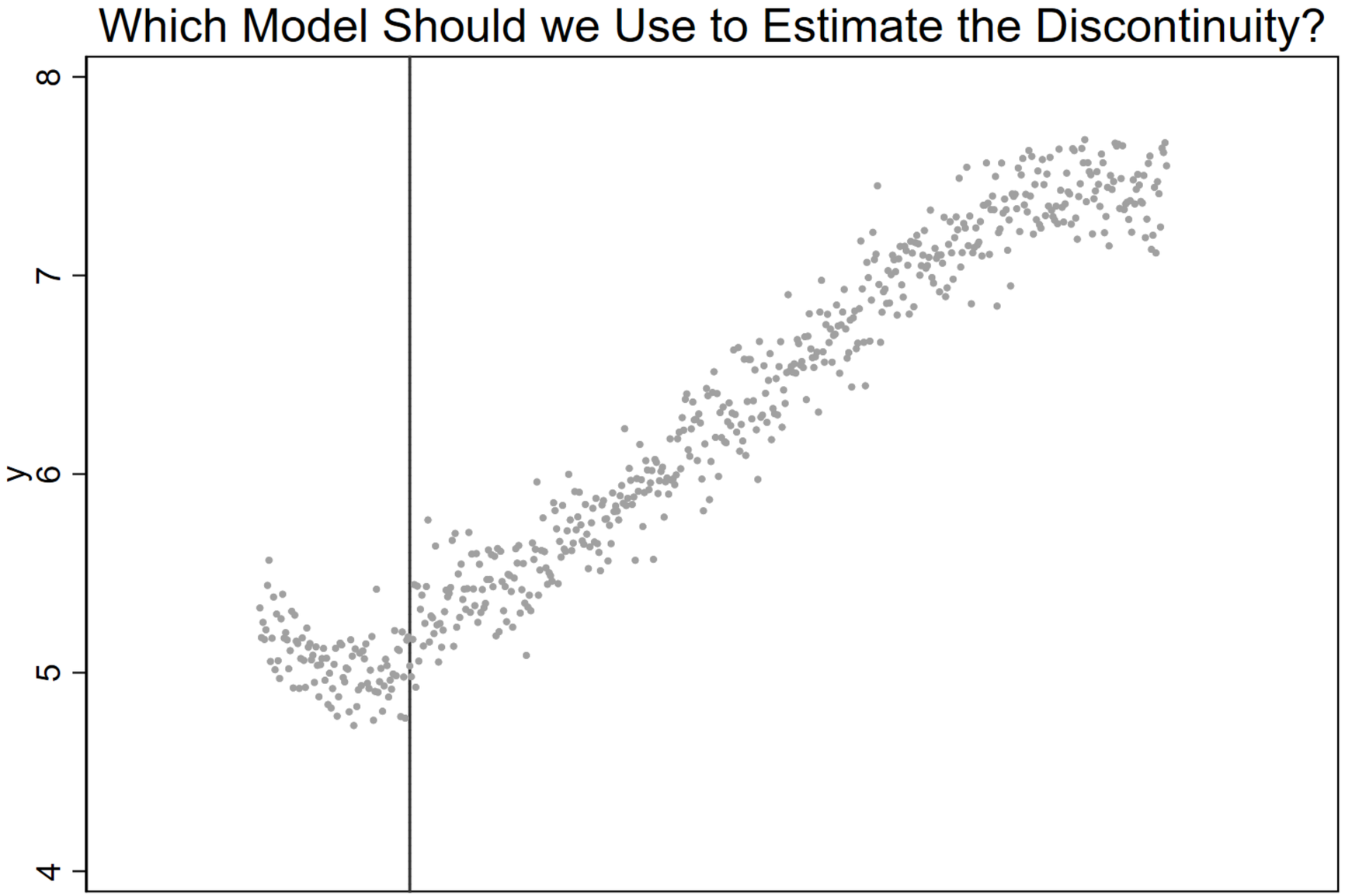
RDD Model Selection

- RDD regarded as a leading quasi-experimental research design
- However, it involves numerous researcher choices e.g. bandwidth, polynomial, kernel, controls etc.
- How to select among the multitudes of potential estimators?
  - Imbens and Kalyanaraman (IK) (2012) and Calonico et al. (CCT) (2014) propose algorithms for BW selection that minimise AMSE of the boundary estimator
  - Pei et al (2021) make a similar suggestion for polynomial order.

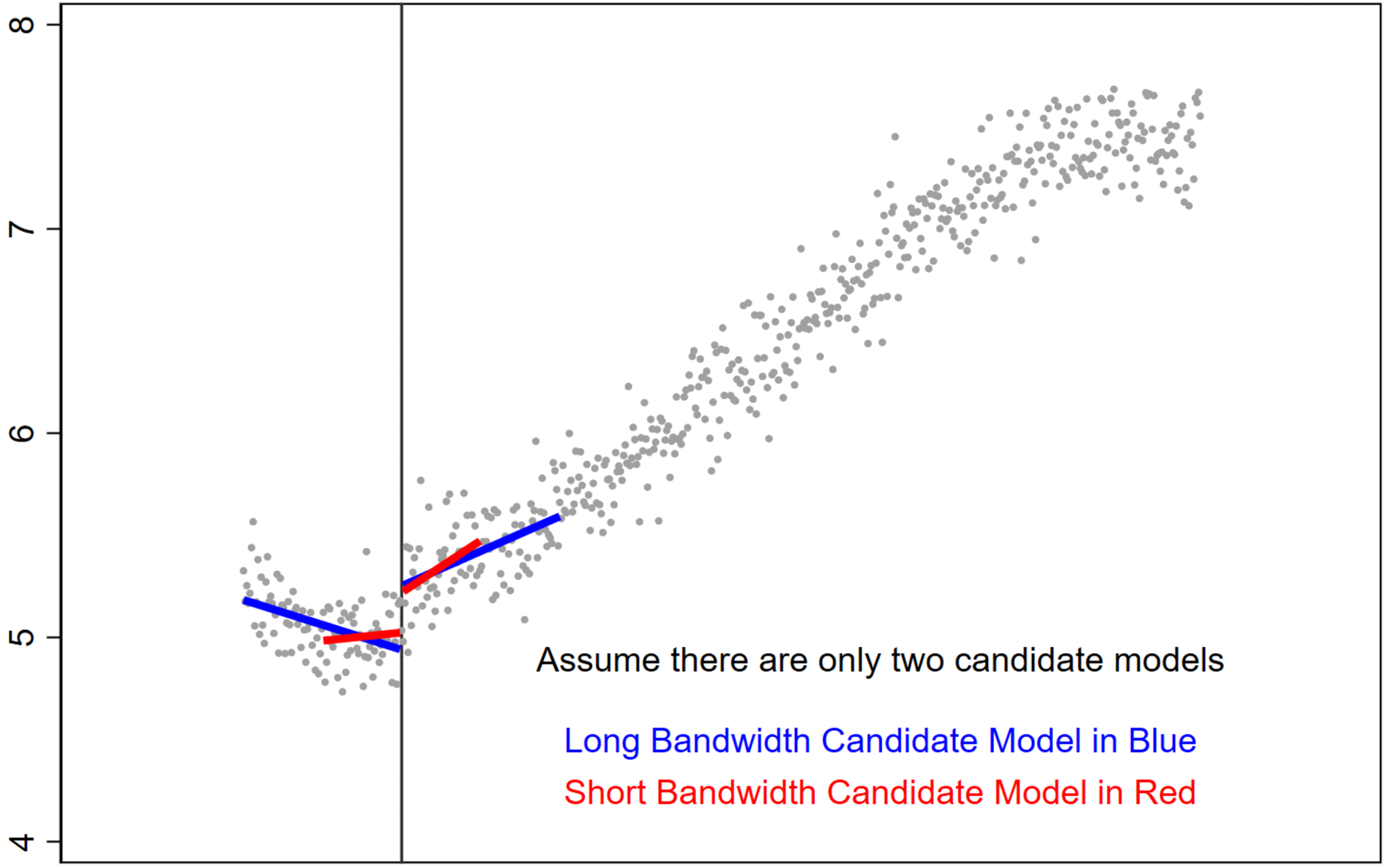
Issues with Model Selection Algorithms

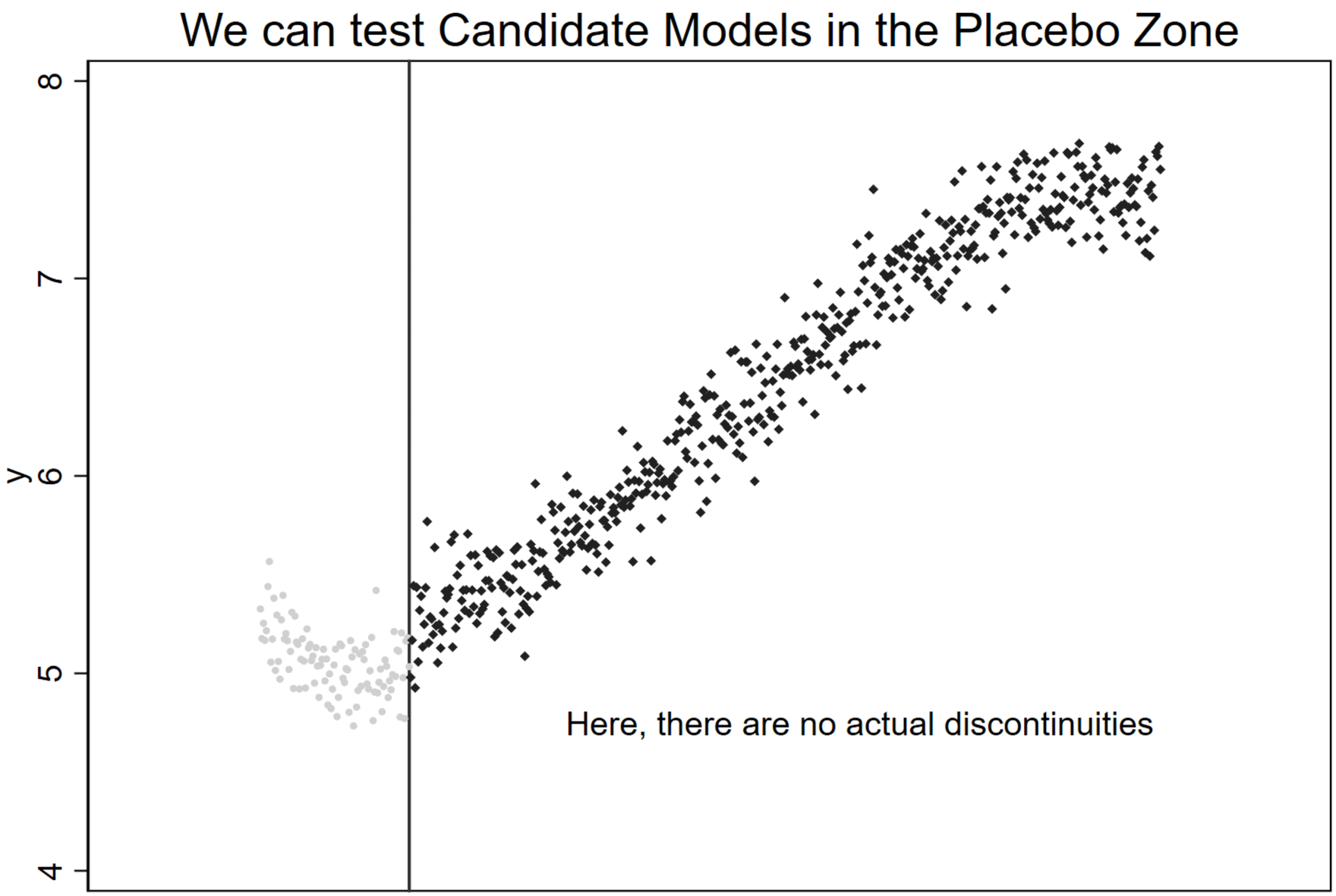
- Existing approaches deal with one choice and hold others constant.
- CCT focussed on inference, not on estimation
- IK/CCT can do poorly in simulations with realistic DGPs (Card et al, 2017).
- While IK/CCT are popular for BW selection, there is no consensus and researchers tend to rely on robustness testing. This may be overly punitive to particular DGPs.

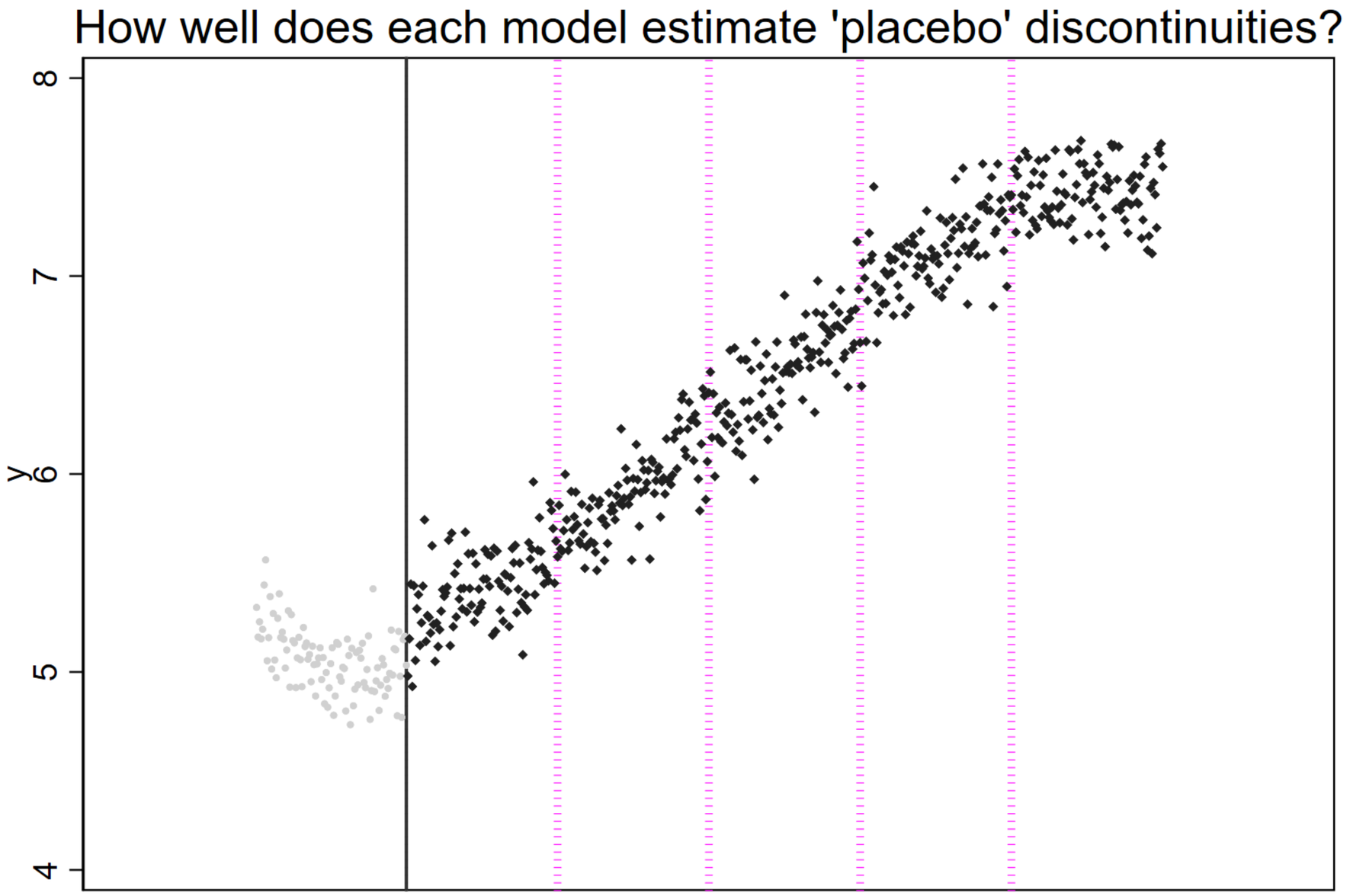
Illustration of our proposal

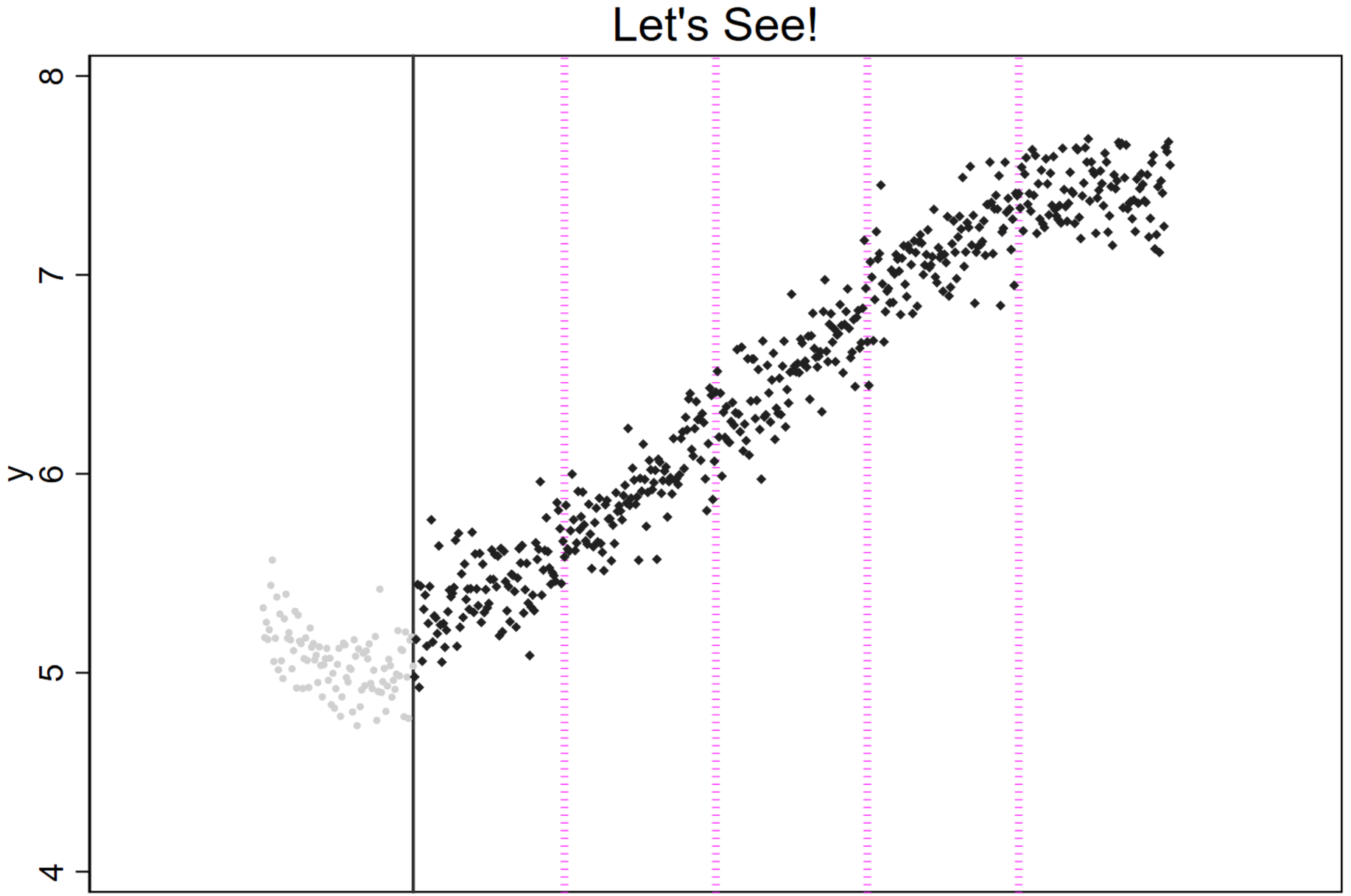


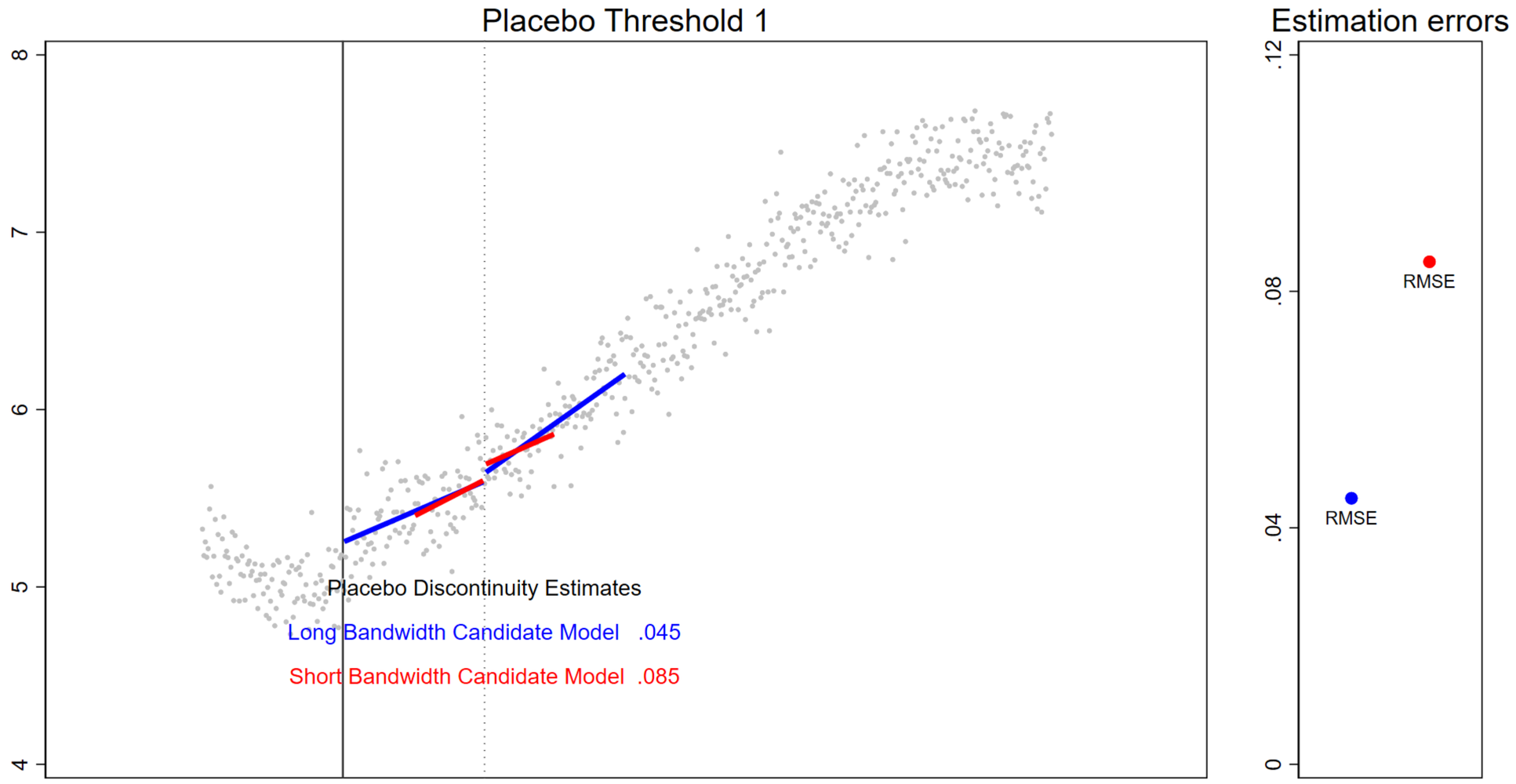
Which Model Should we Use to Estimate the Discontinuity?

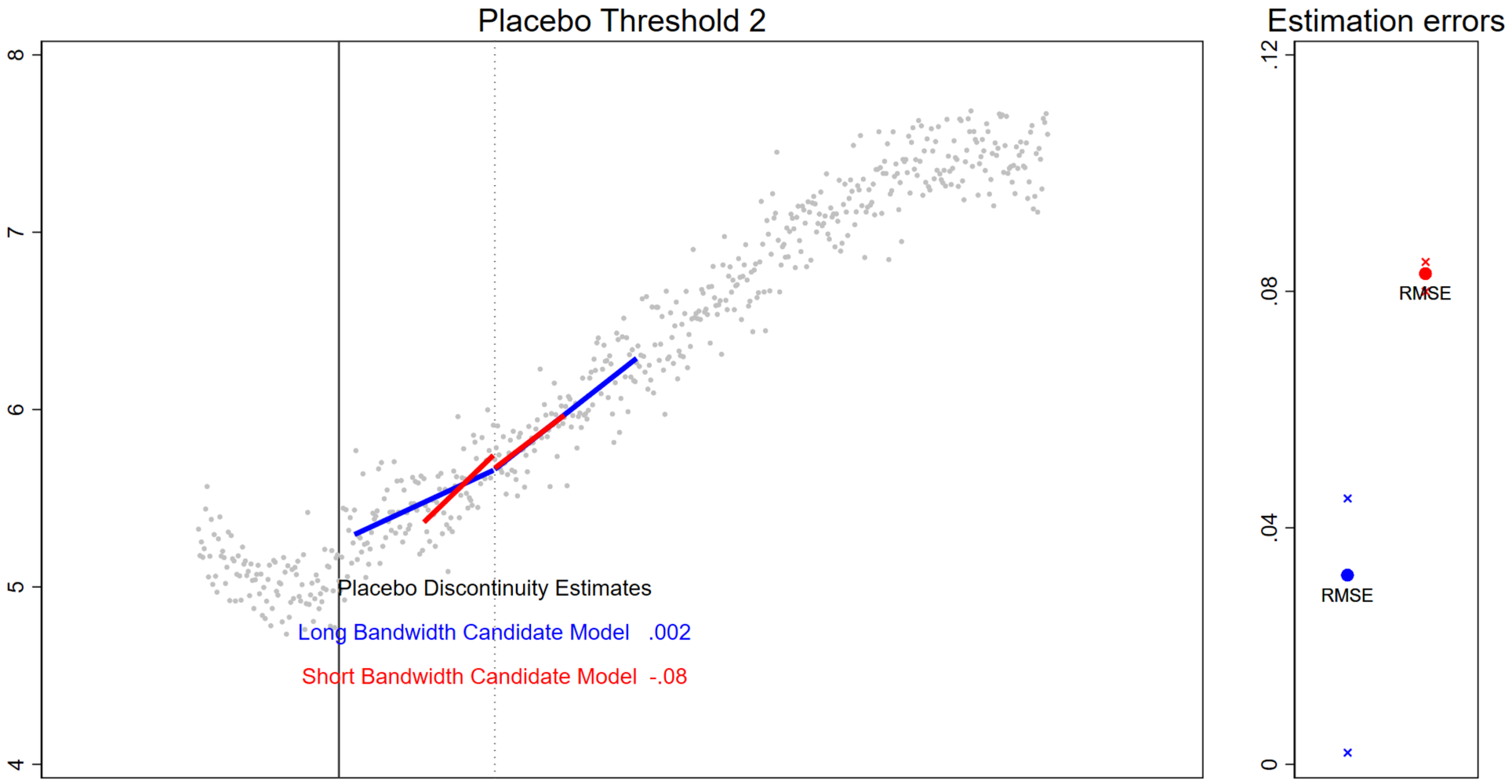


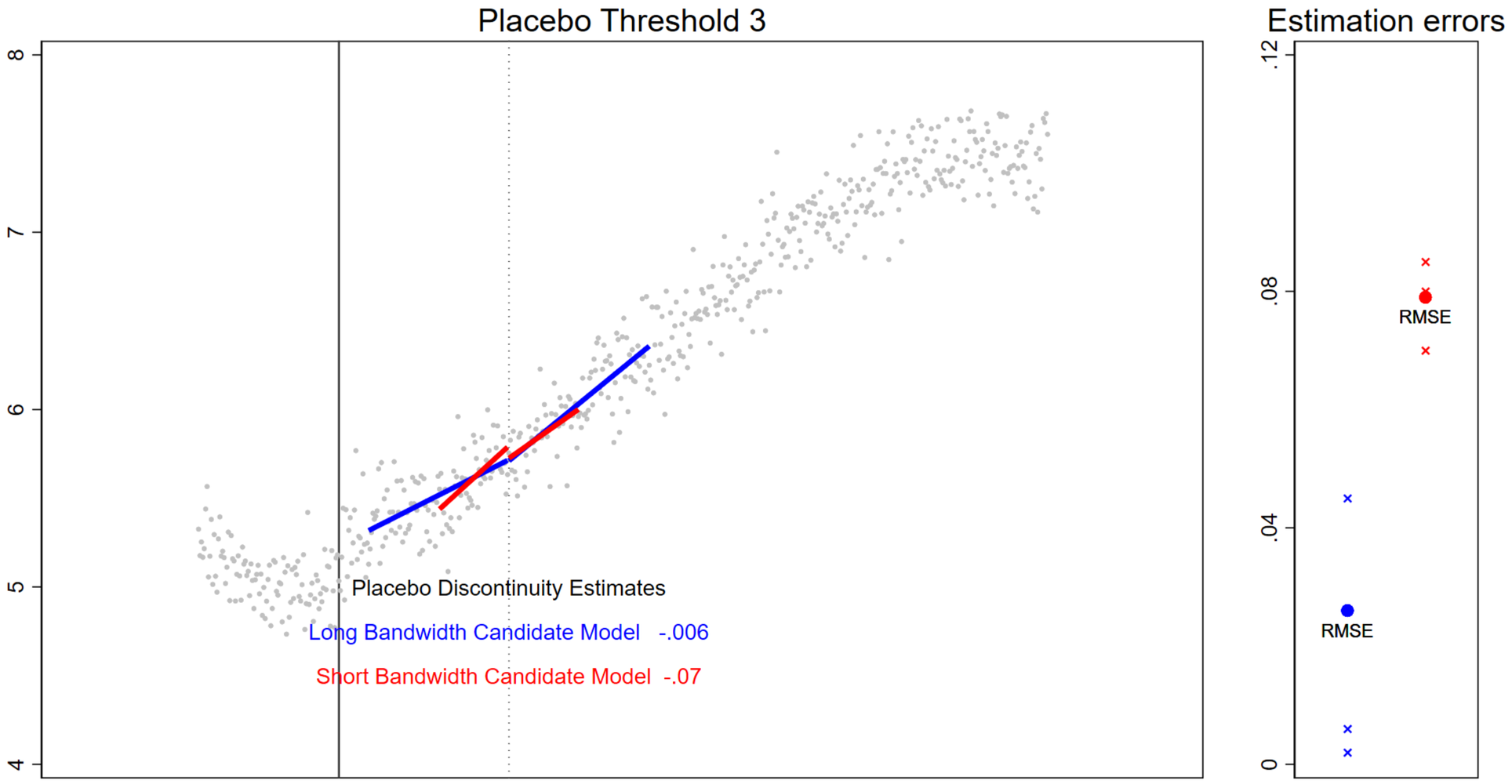

We can test Candidate Models in the Placebo Zone
y
8
7
6
5
4
Here, there are no actual discontinuities

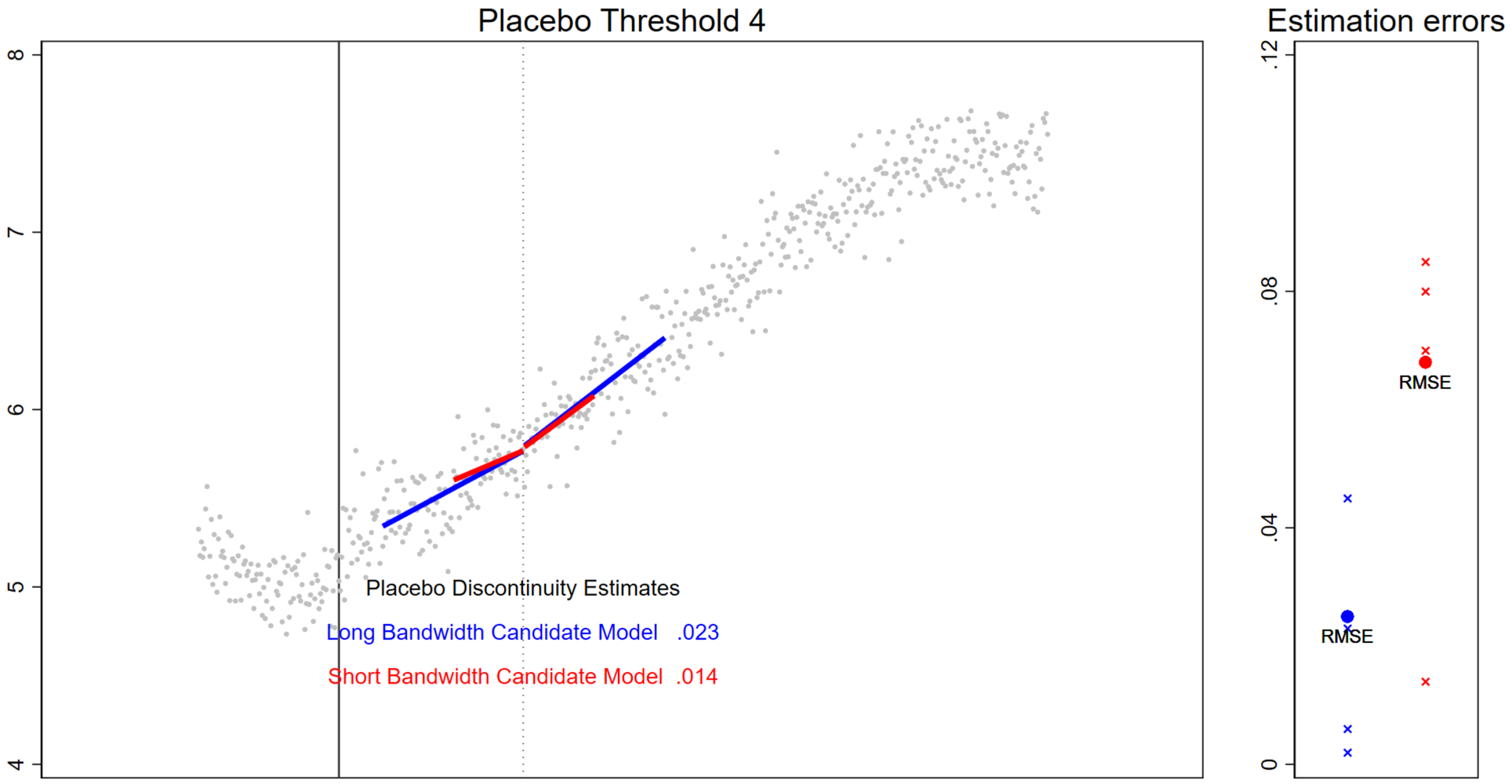

How well does each model estimate 'placebo' discontinuities?
y
4
5
6
7
8

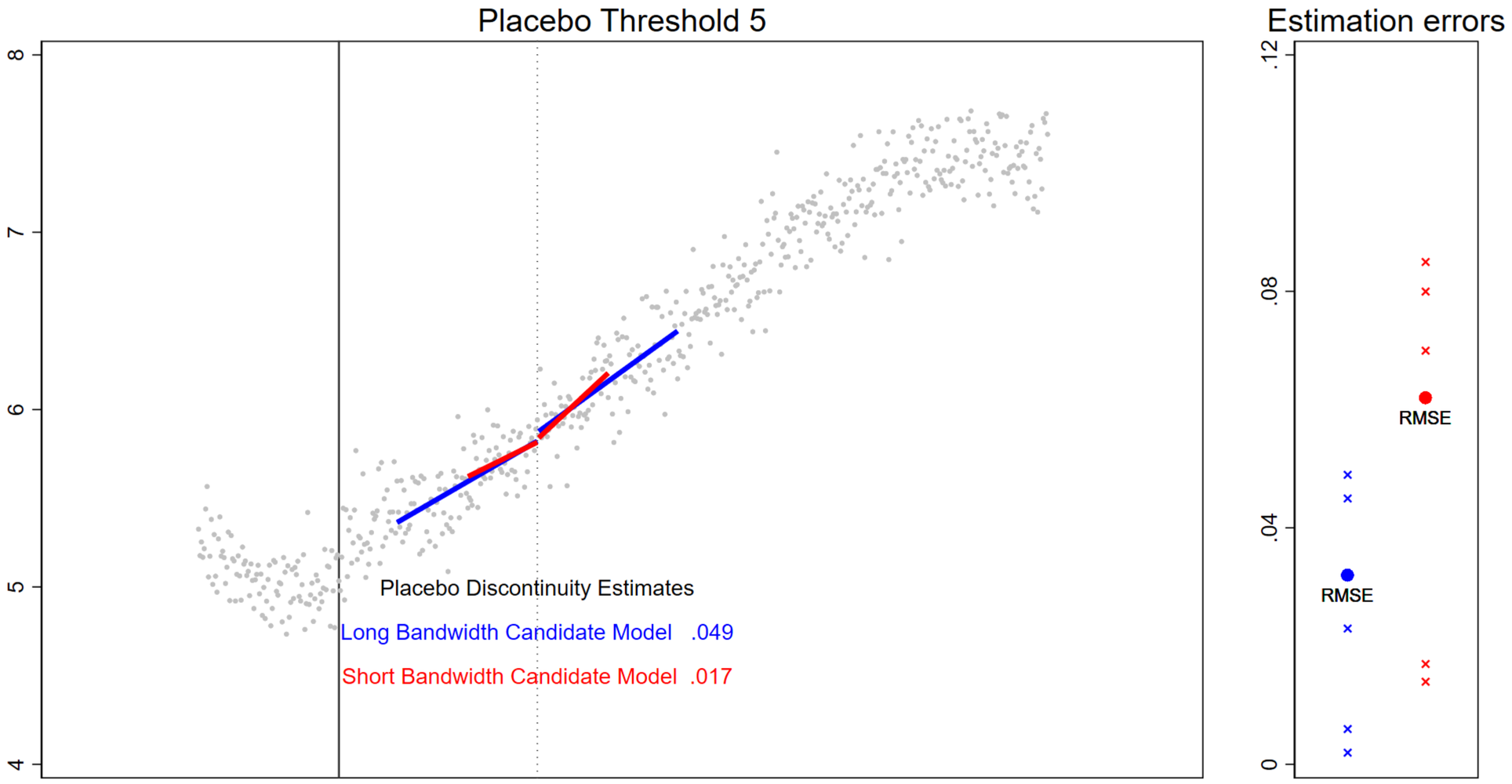

Let's See!
y
8
7
6
5
4

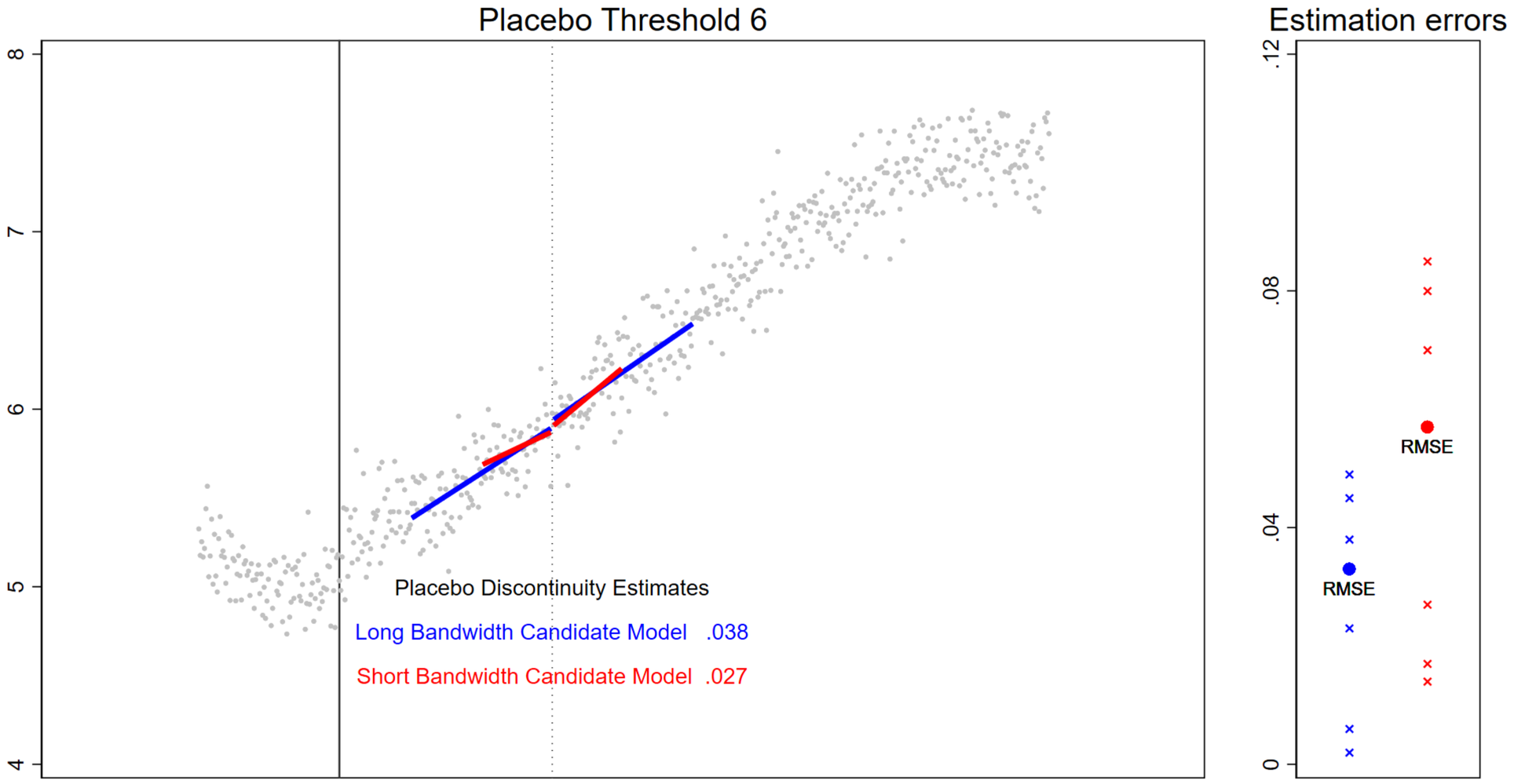

Placebo Threshold 1
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .045
Short Bandwidth Candidate Model .085
Estimation errors
.12
.08
.04
0
RMSE
RMSE

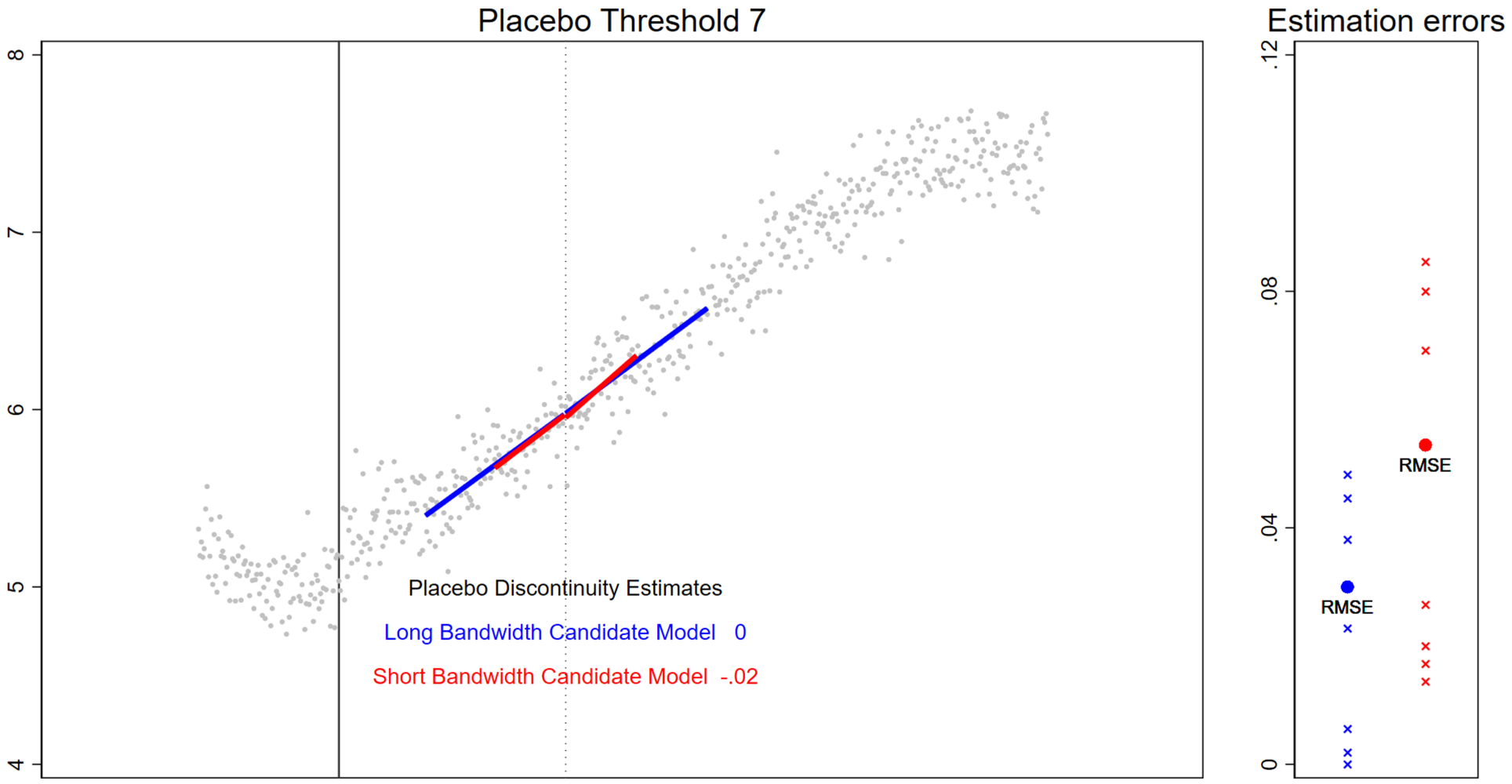

Placebo Threshold 2
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .002
Short Bandwidth Candidate Model -.08
Estimation errors
.12
.08
.04
0
RMSE
RMSE

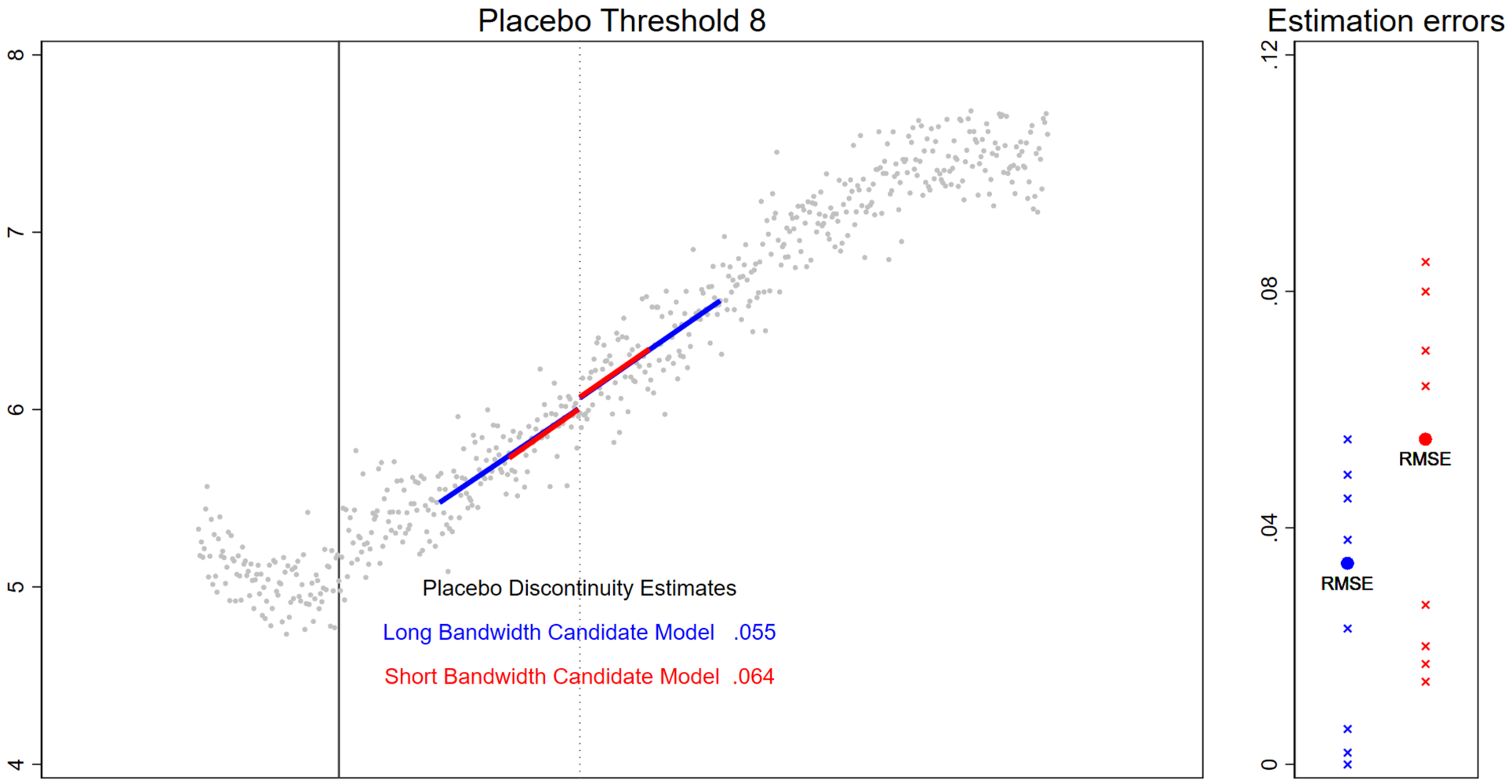

Placebo Threshold 3
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.006
Short Bandwidth Candidate Model -.07
Estimation errors
.12
.08
.04
0
RMSE
RMSE

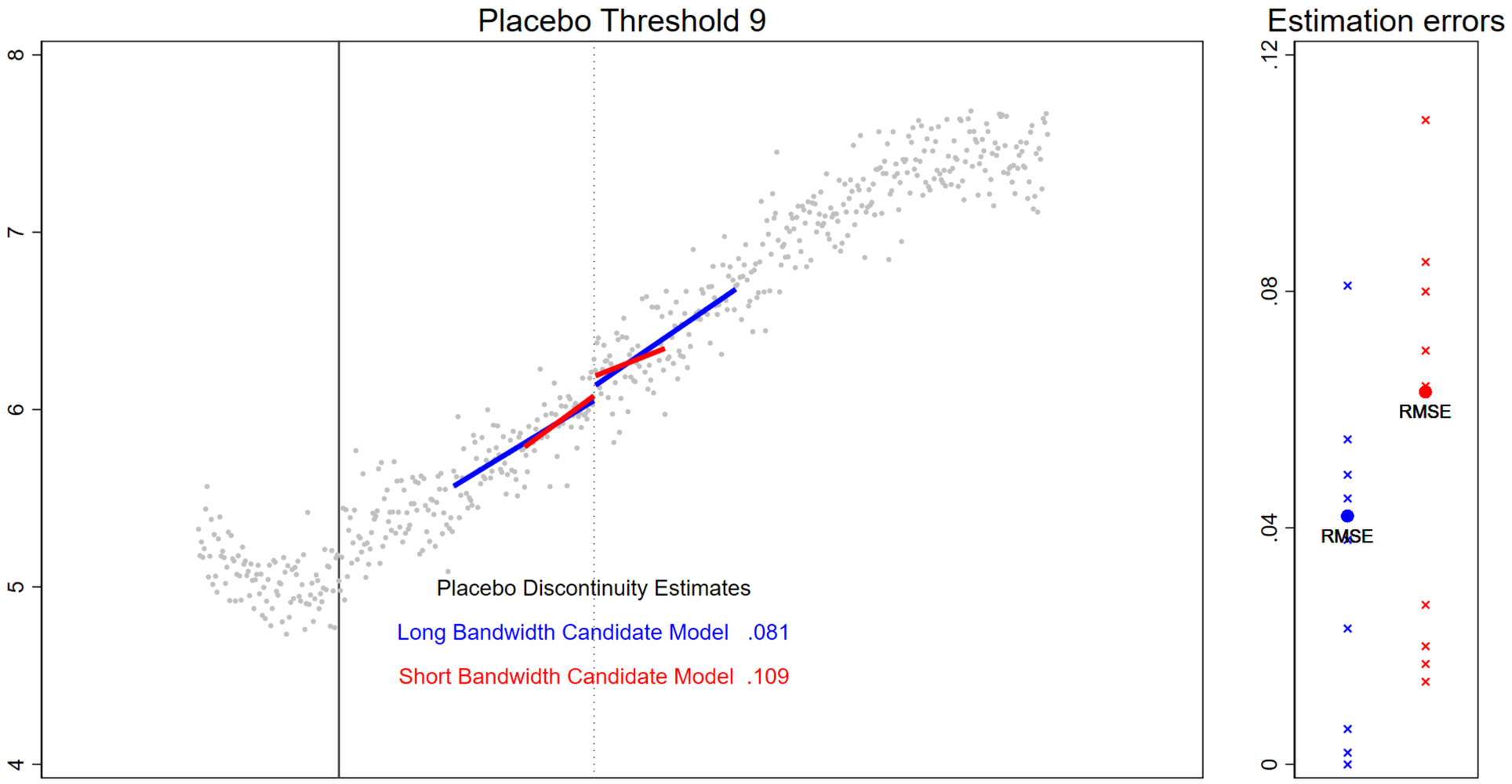

Placebo Threshold 4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .023
Short Bandwidth Candidate Model .014
8
7
6
5
4
Estimation errors
.12
.08
.04
0
RMSE
RMSE

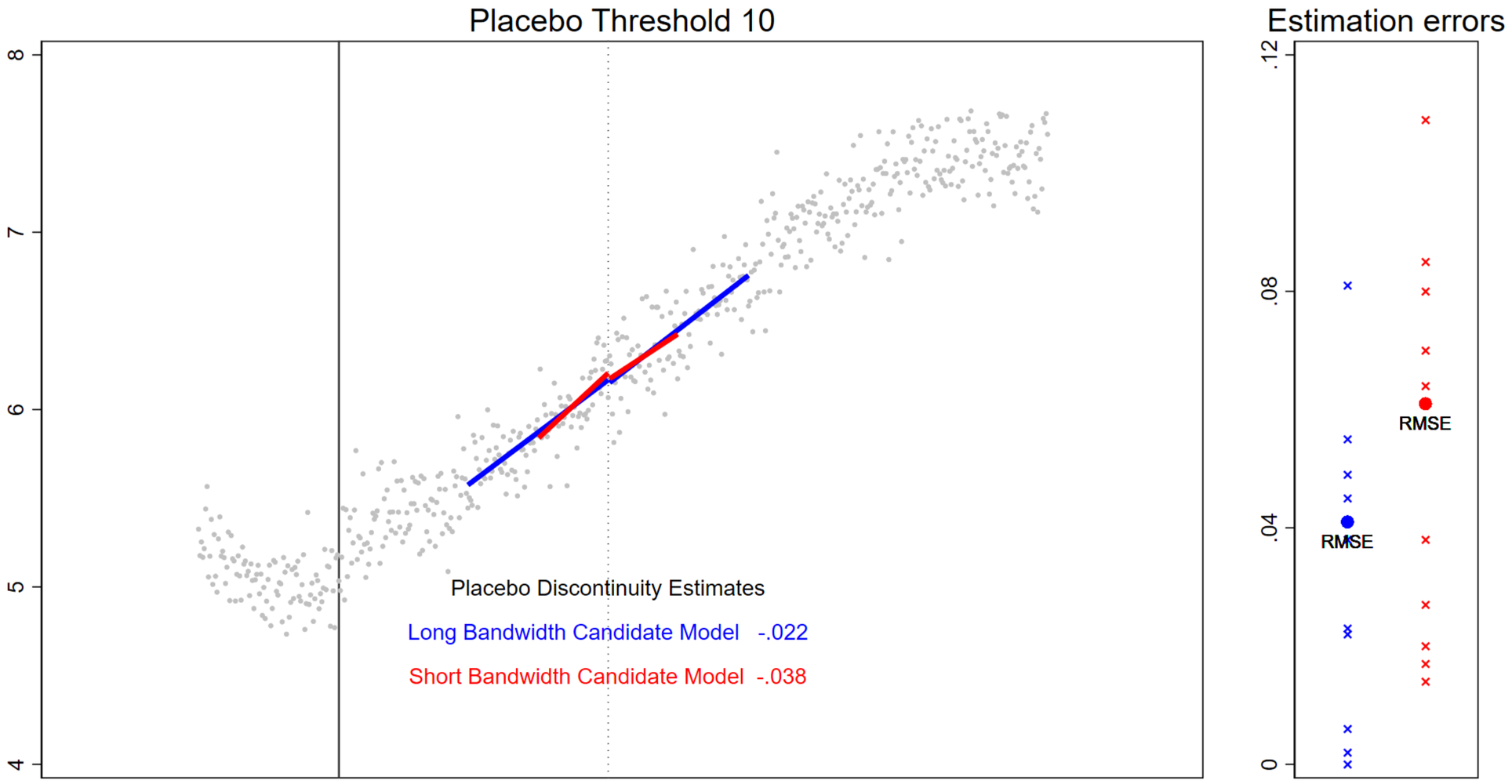

Placebo Threshold 5
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .049
Short Bandwidth Candidate Model .017
Estimation errors
.12
.08
.04
0
RMSE
RMSE

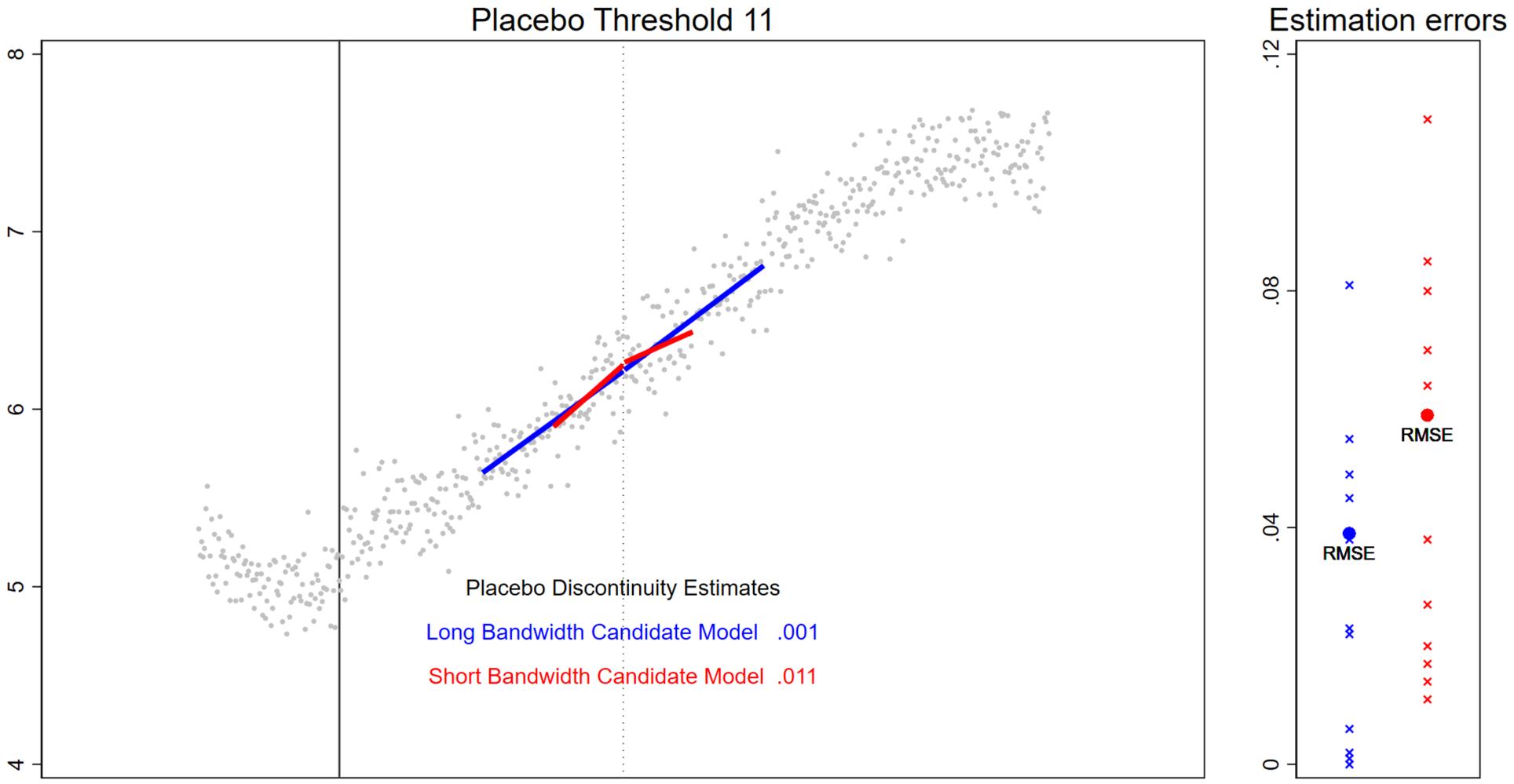

Placebo Threshold 6
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .038
Short Bandwidth Candidate Model .027
Estimation errors
.12
.08
.04
0
RMSE
RMSE

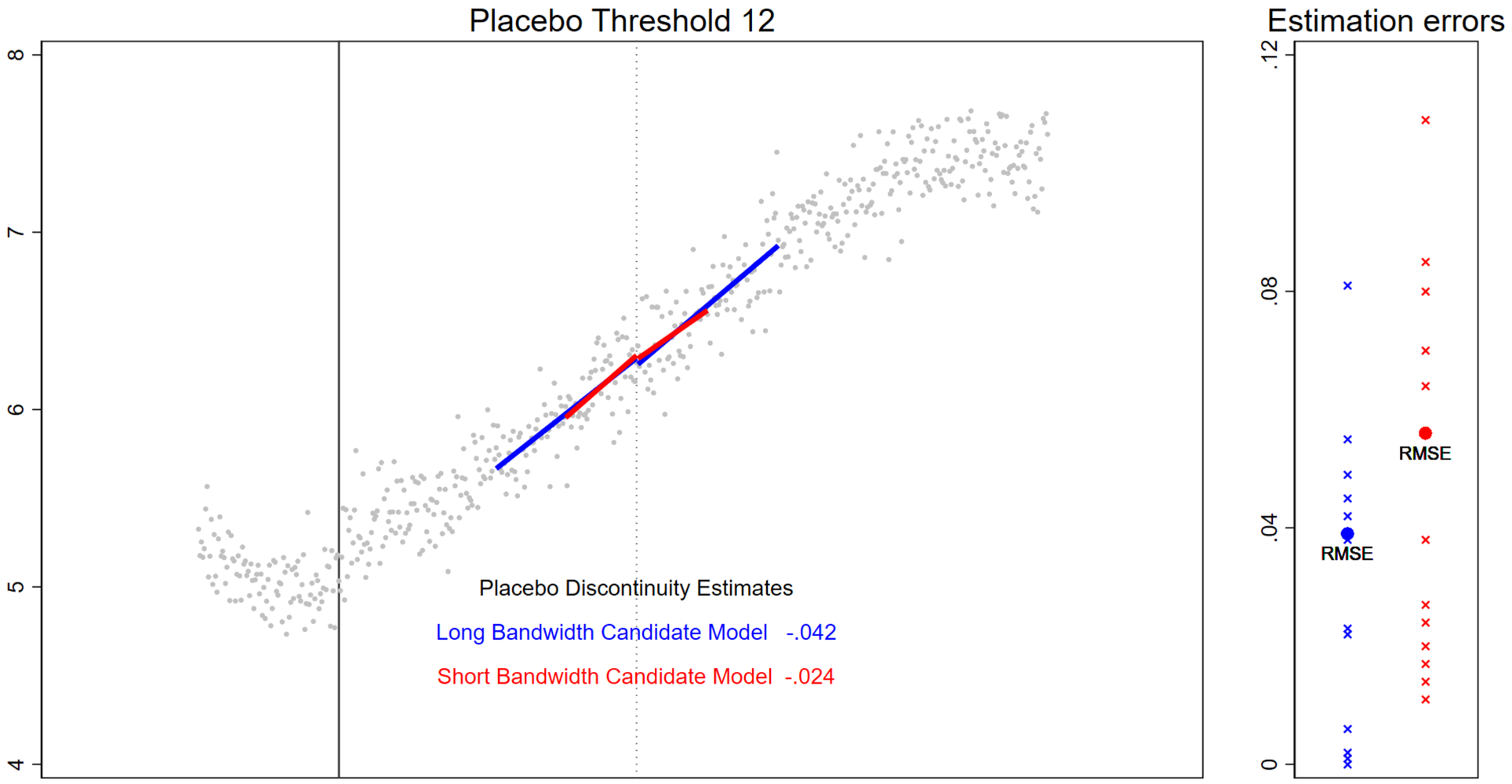

Placebo Threshold 7
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model 0
Short Bandwidth Candidate Model -.02
Estimation errors
.12
.08
.04
0
RMSE
RMSE

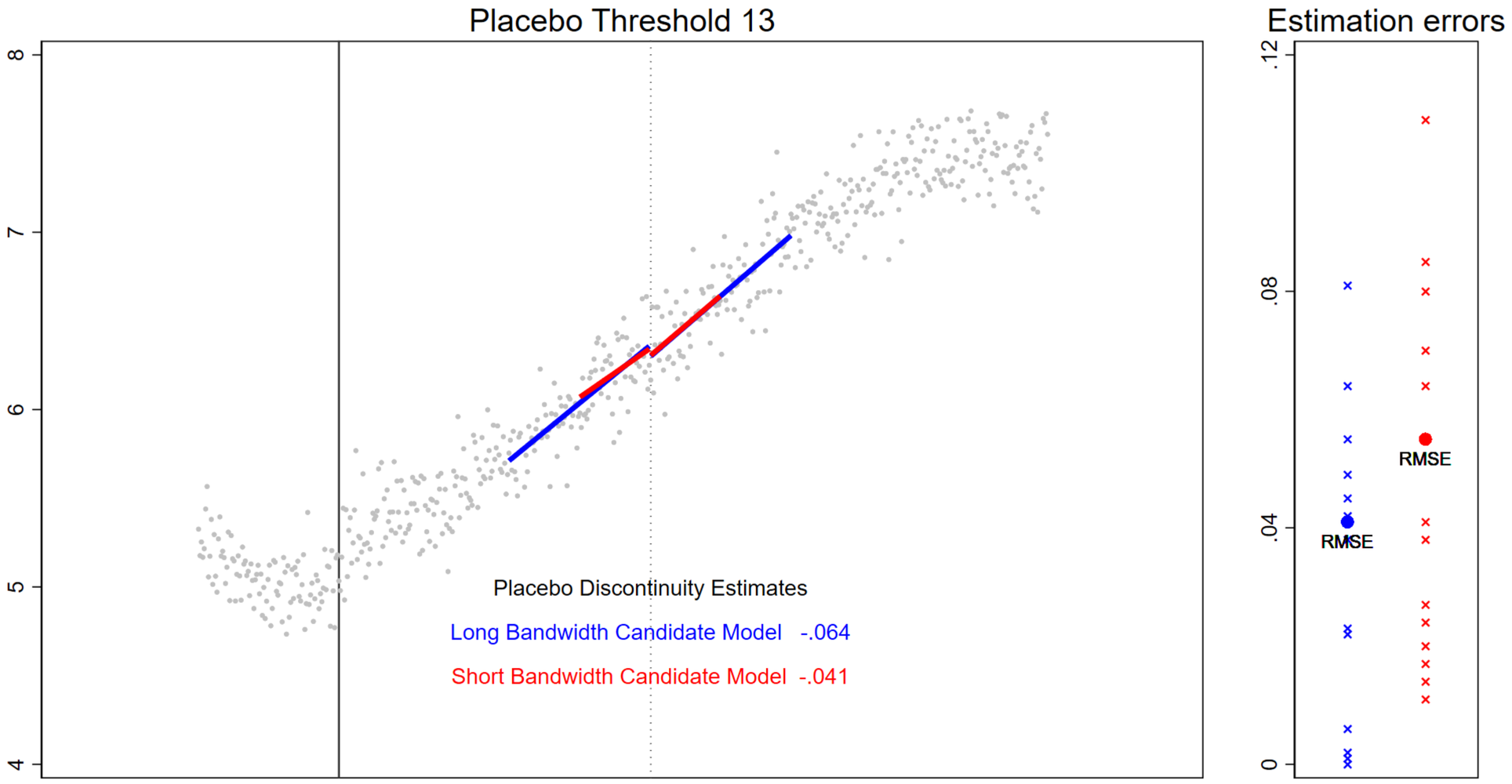

Placebo Threshold 8
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .055
Short Bandwidth Candidate Model .064
Estimation errors
.12
.08
.04
0
RMSE
RMSE

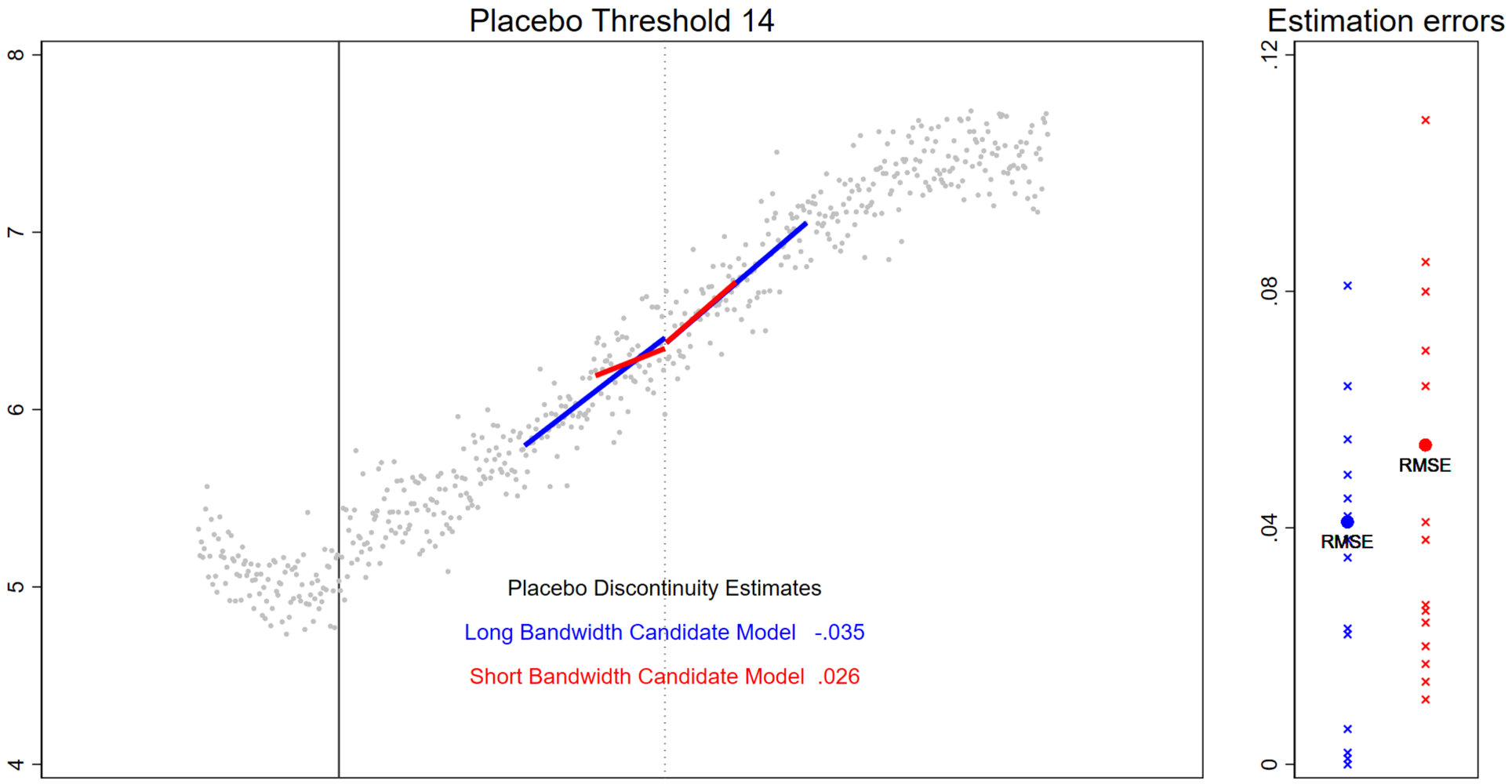

Placebo Threshold 9
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .081
Short Bandwidth Candidate Model .109
Estimation errors
.12
.08
.04
0
RMSE
RMSE

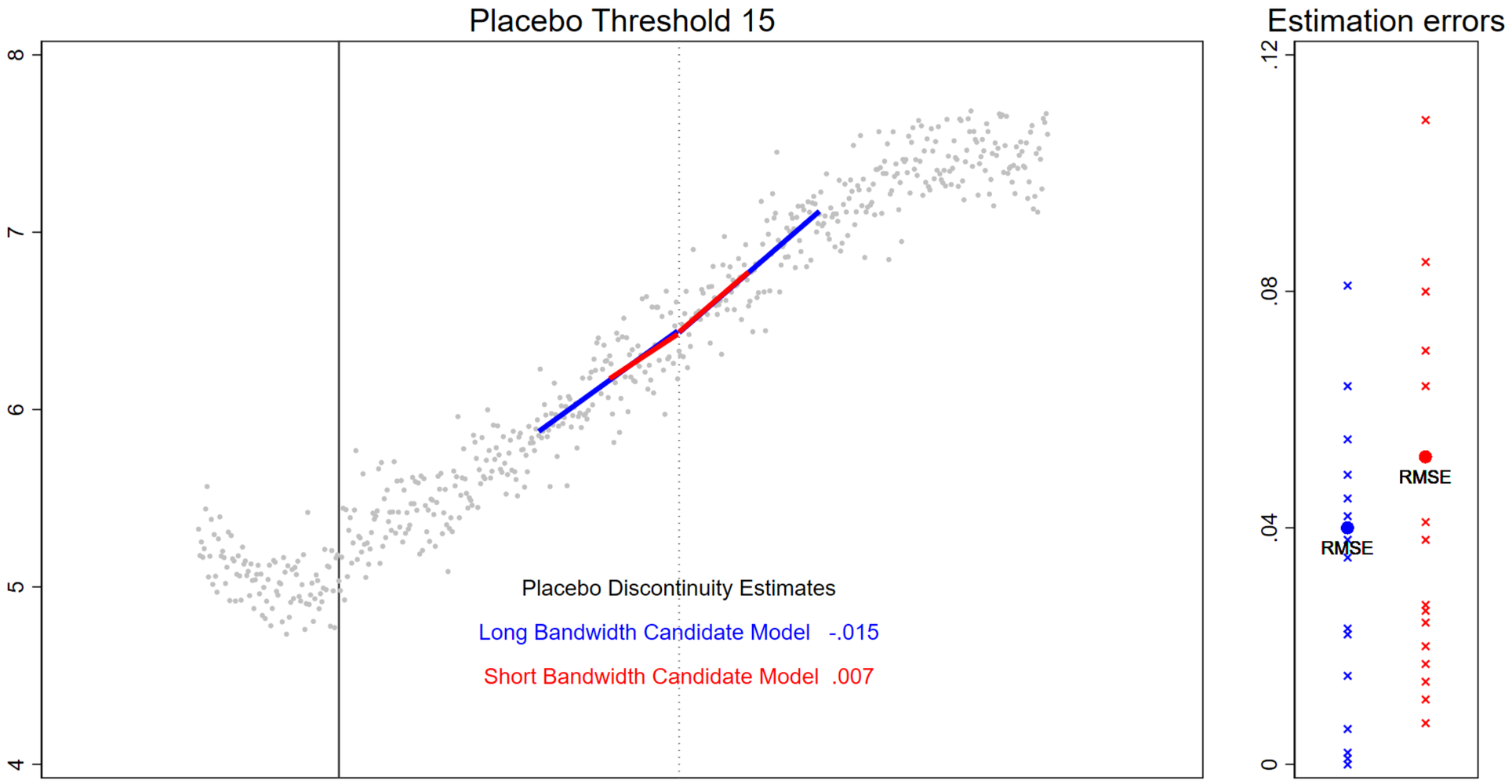

Placebo Threshold 10
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.022
Short Bandwidth Candidate Model -.038
Estimation errors
.12
.08
.04
0
RMSE
RMSE

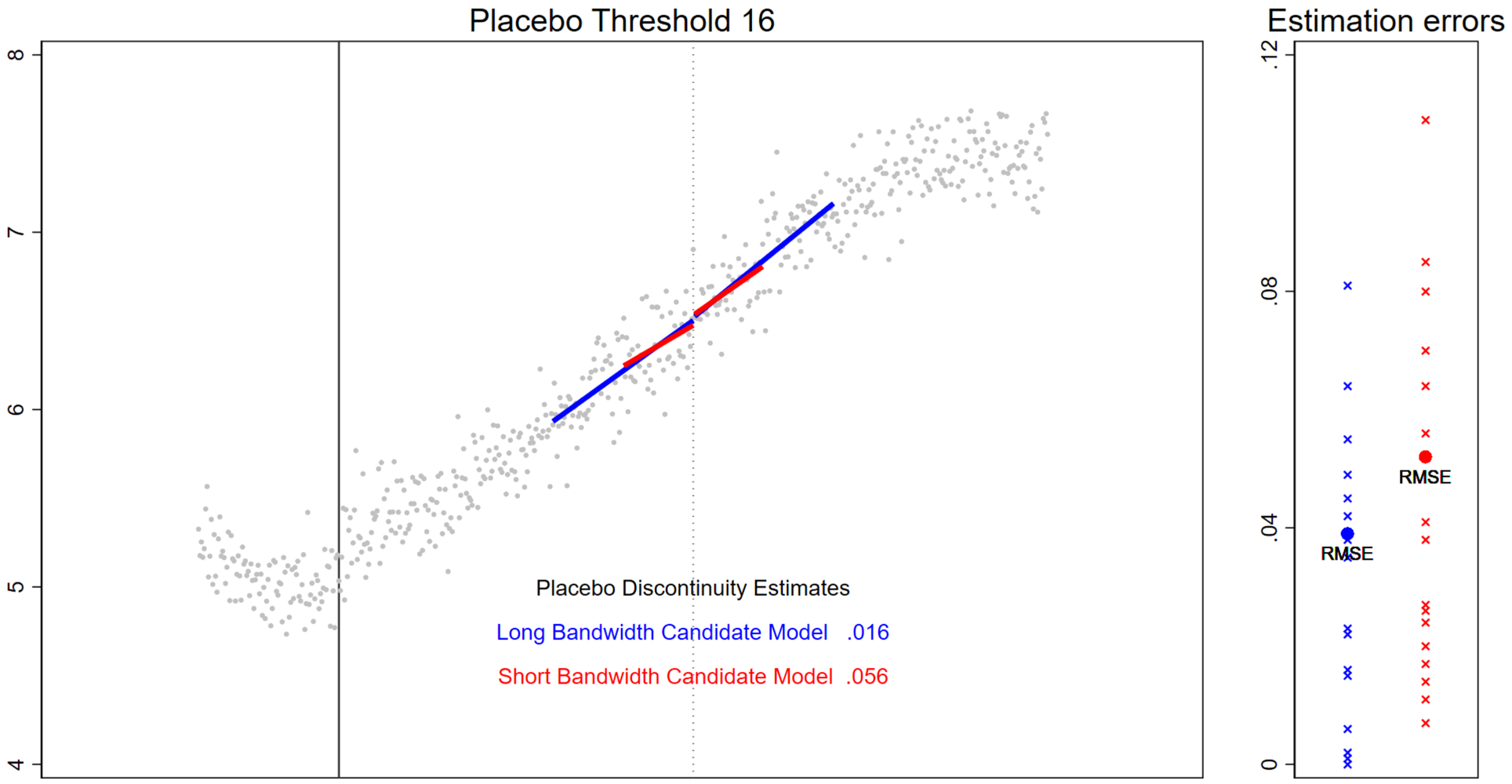

Placebo Threshold 11
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .001
Short Bandwidth Candidate Model .011
Estimation errors
.12
.08
.04
0
RMSE
RMSE

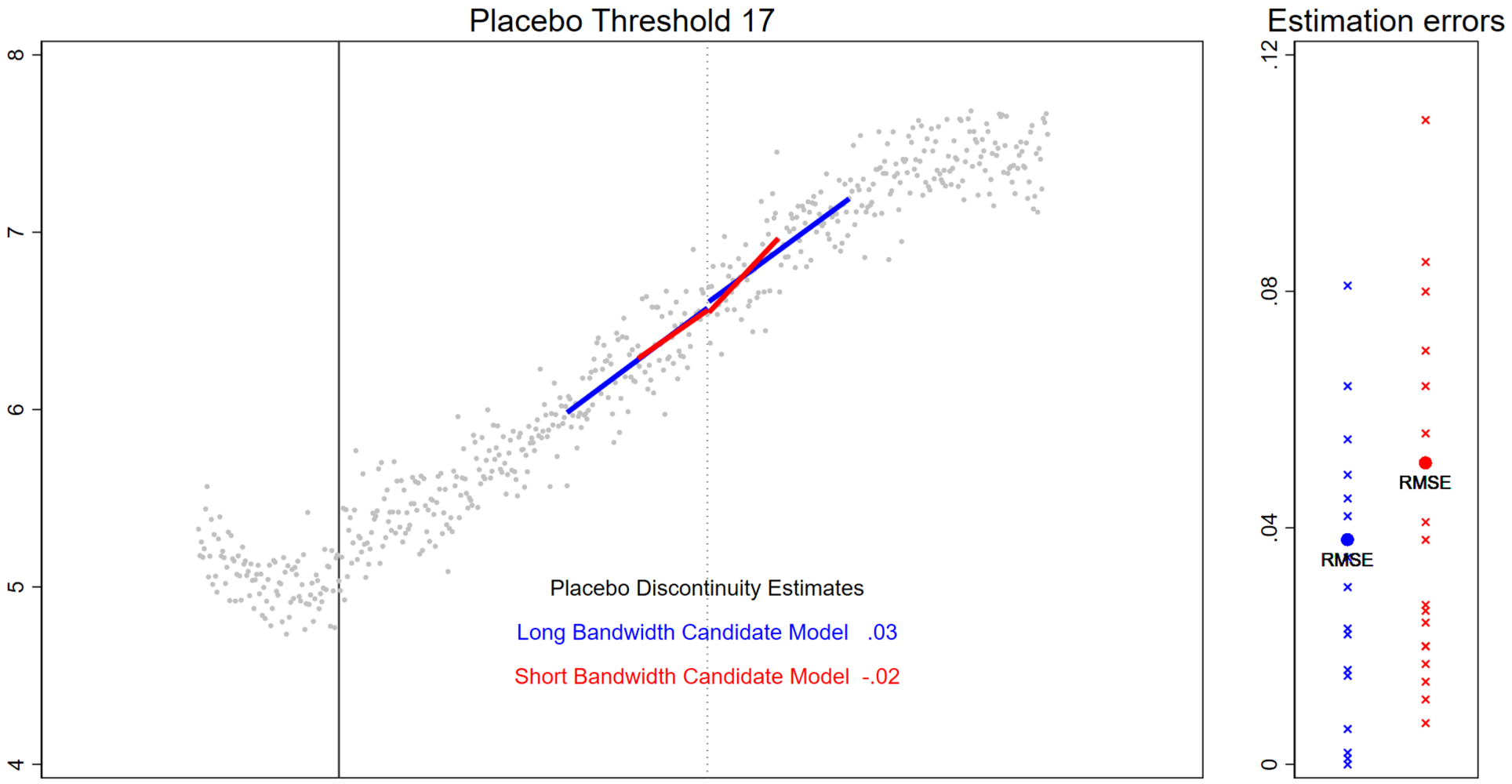

Placebo Threshold 12
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.042
Short Bandwidth Candidate Model -.024
Estimation errors
.12
.08
.04
0
RMSE
RMSE

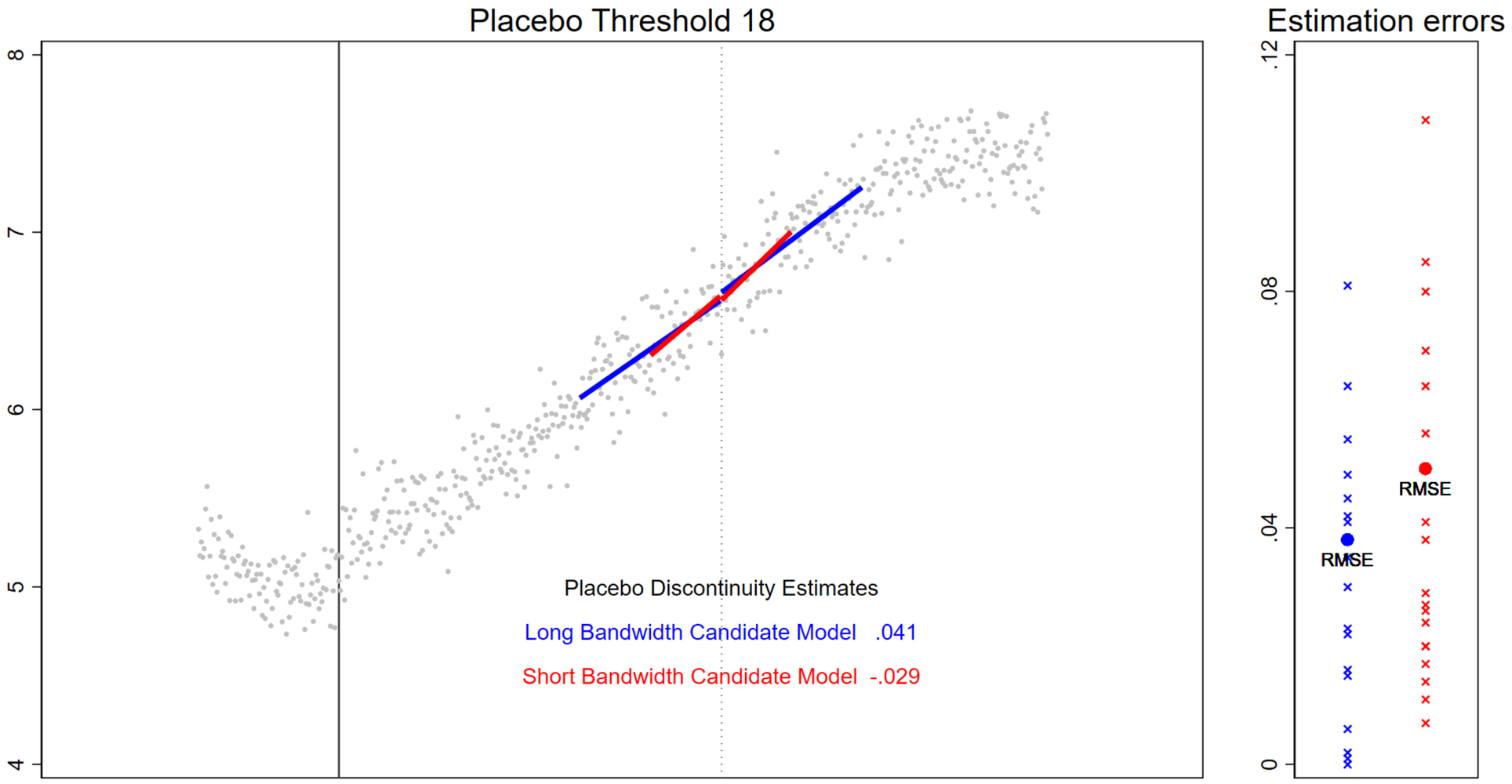

Placebo Threshold 13
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.064
Short Bandwidth Candidate Model -.041
Estimation errors
.12
.08
.04
0
RMSE
RMSE

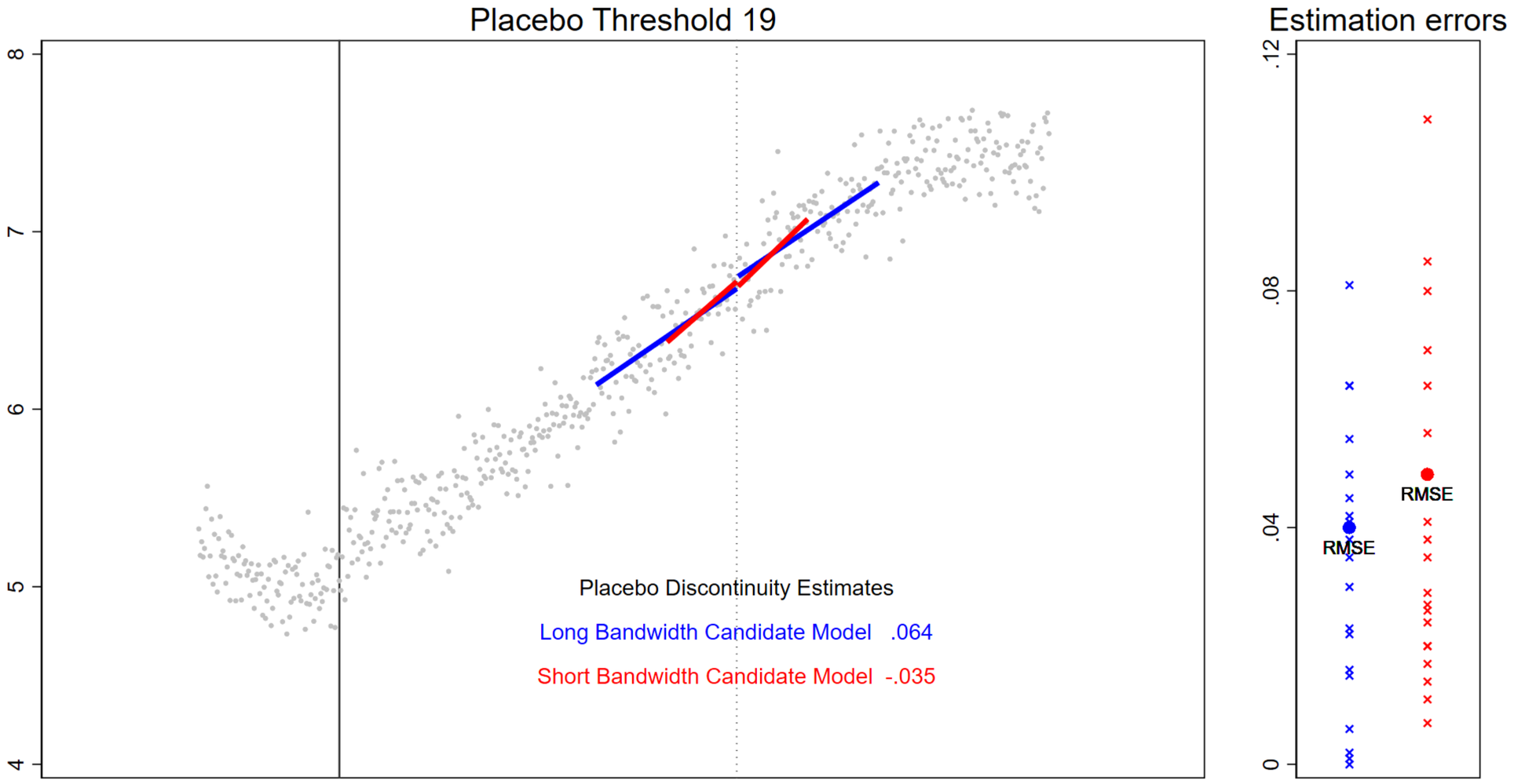

Placebo Threshold 14
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.035
Short Bandwidth Candidate Model .026
Estimation errors
.12
.08
.04
0
RMSE
RMSE

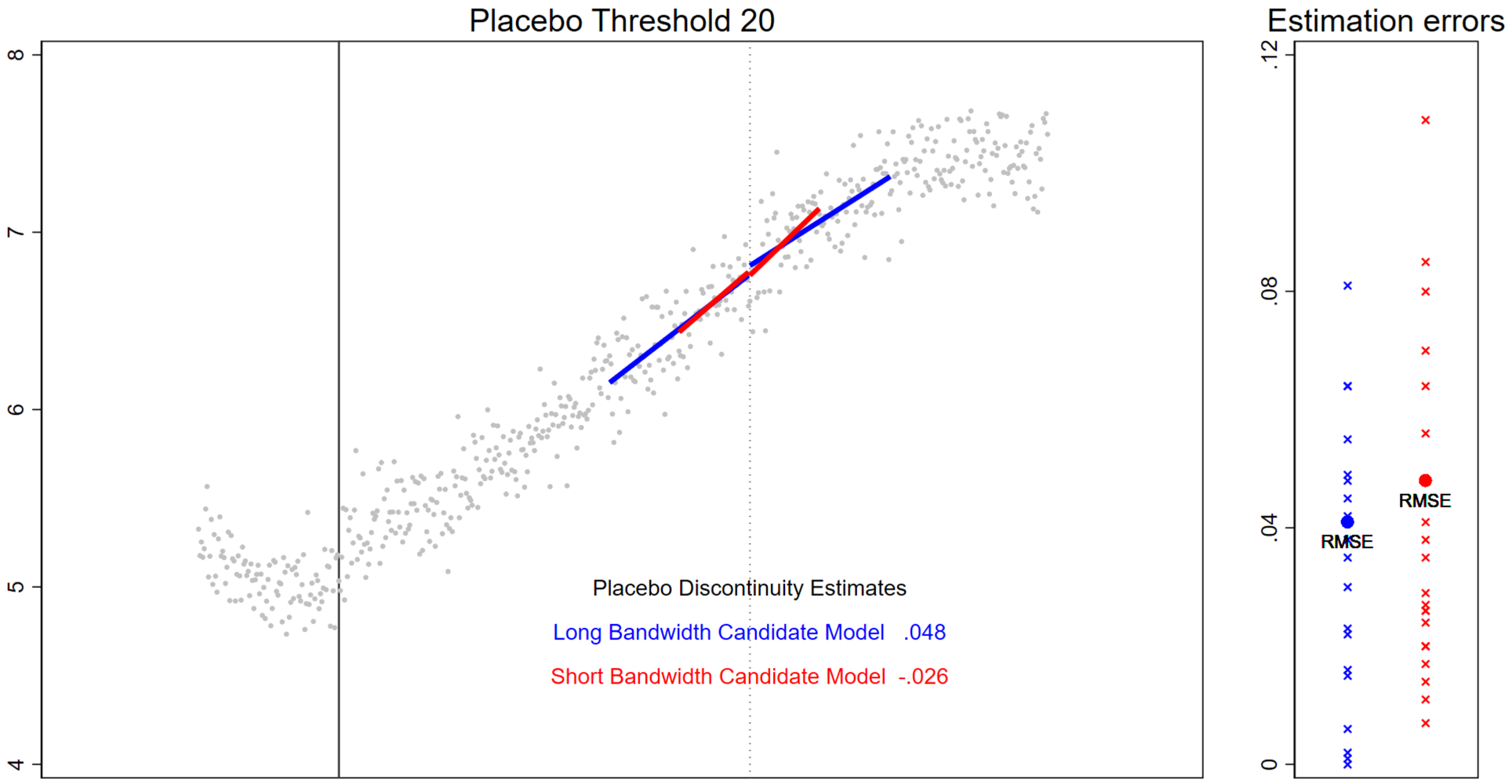

Placebo Threshold 15
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.015
Short Bandwidth Candidate Model .007
Estimation errors
.12
.08
.04
0
RMSE
RMSE

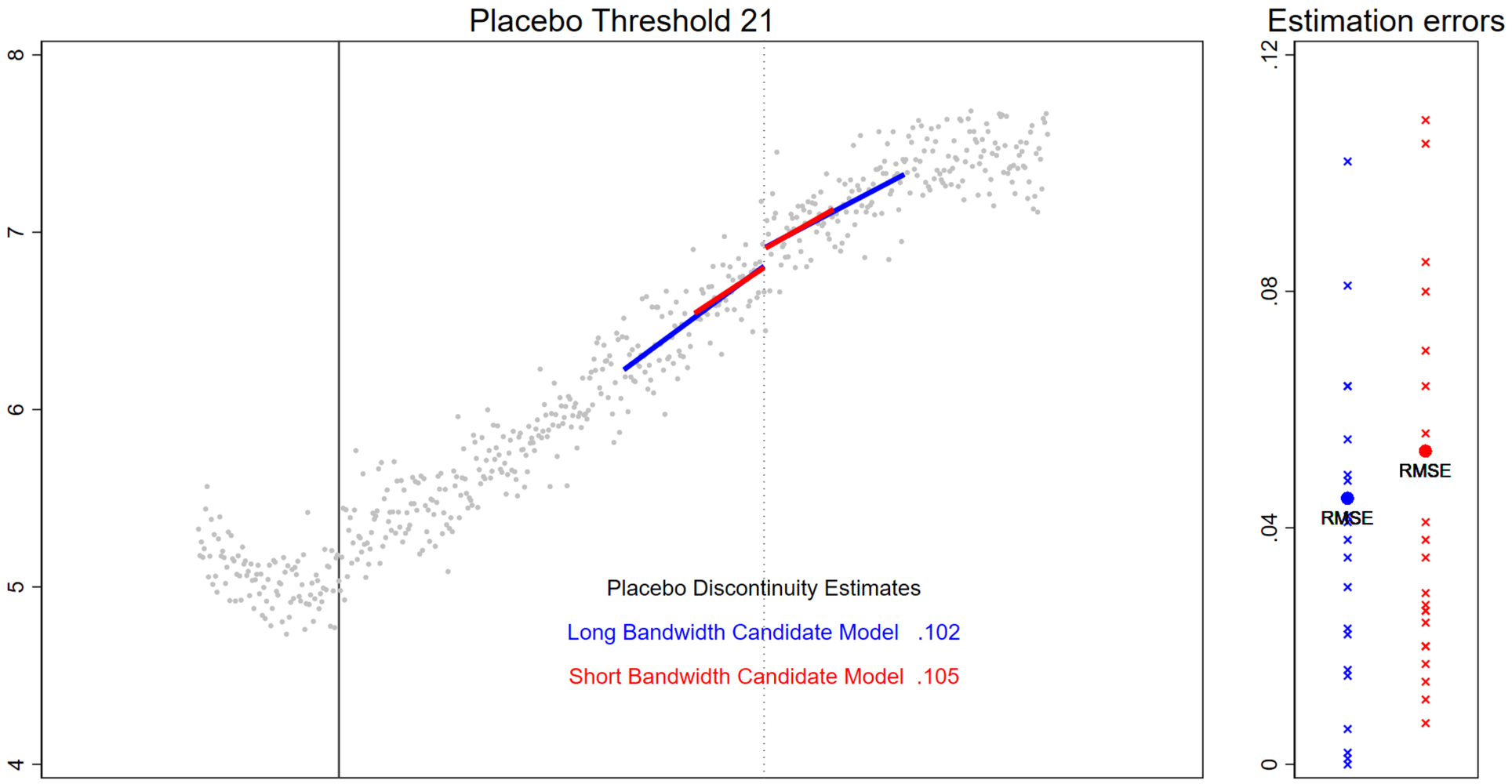

Placebo Threshold 16
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .016
Short Bandwidth Candidate Model .056
Estimation errors
.12
.08
.04
0
RMSE
RMSE

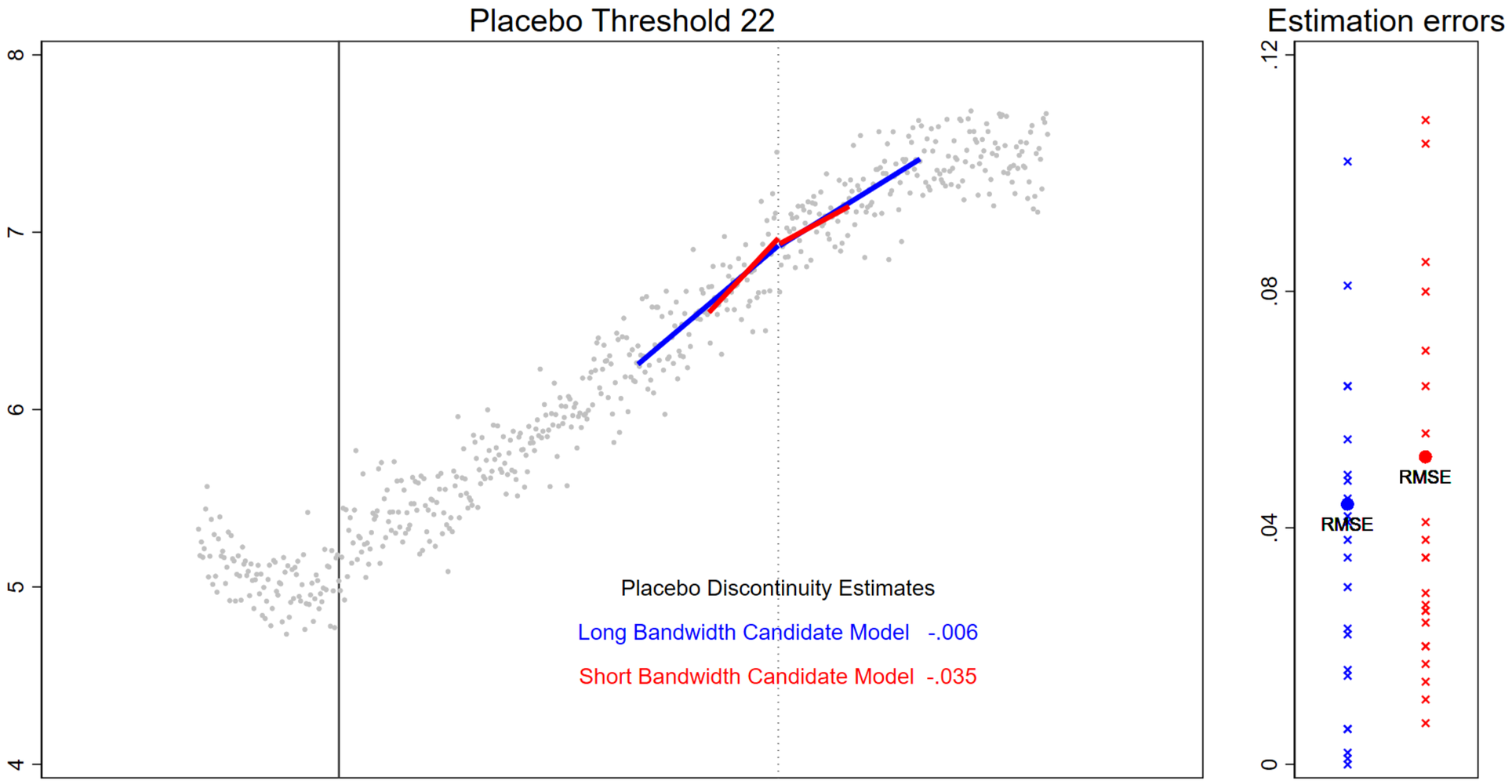

Placebo Threshold 17
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .03
Short Bandwidth Candidate Model -.02
Estimation errors
.12
.08
.04
0
RMSE
RMSE

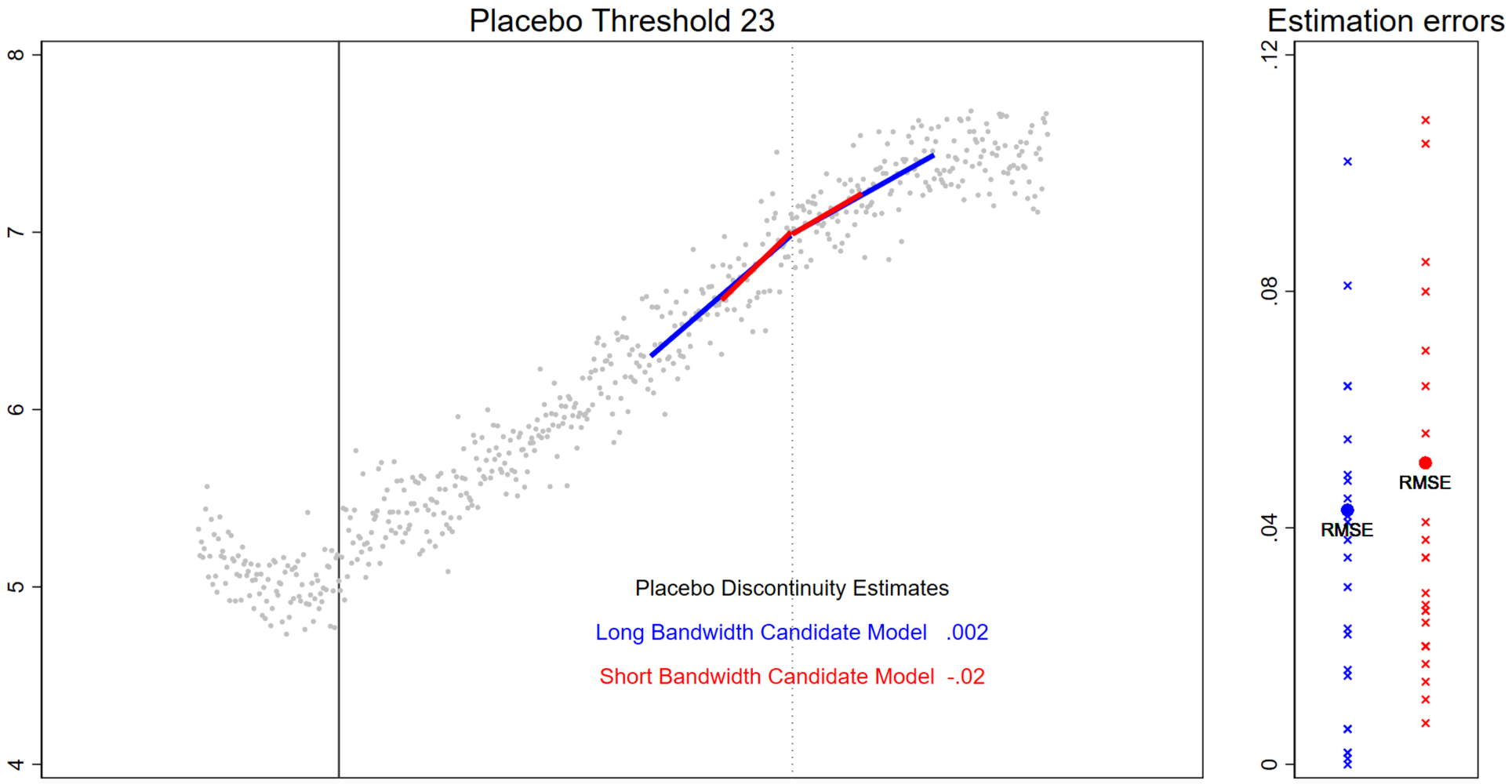

Placebo Threshold 18
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .041
Short Bandwidth Candidate Model -.029
Estimation errors
.12
.08
.04
0
RMSE
RMSE

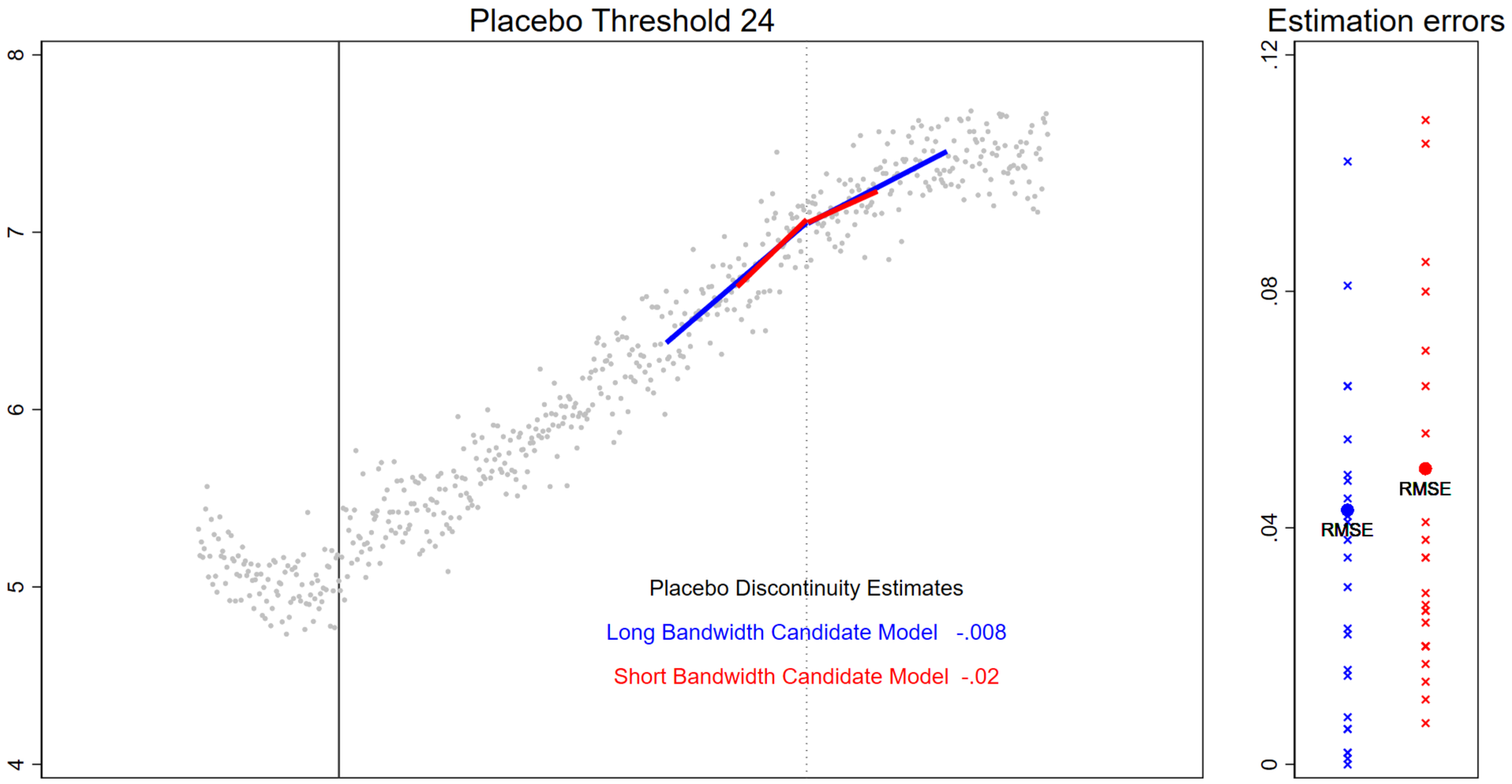

Placebo Threshold 19
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .064
Short Bandwidth Candidate Model -.035
Estimation errors
.12
.08
.04
0
RMSE
RMSE

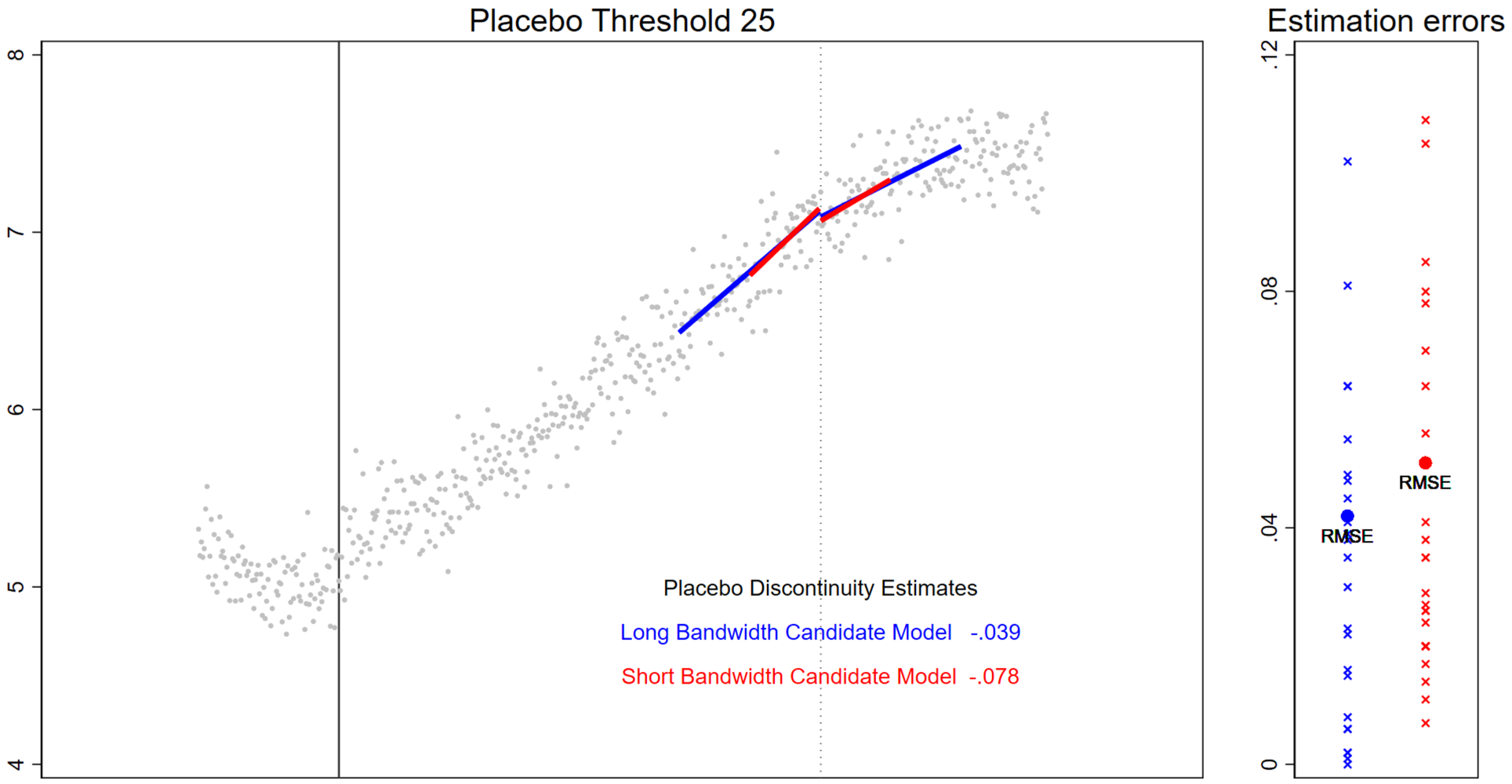

Placebo Threshold 20
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .048
Short Bandwidth Candidate Model -.026
Estimation errors
.12
.08
.04
0
RMSE
RMSE

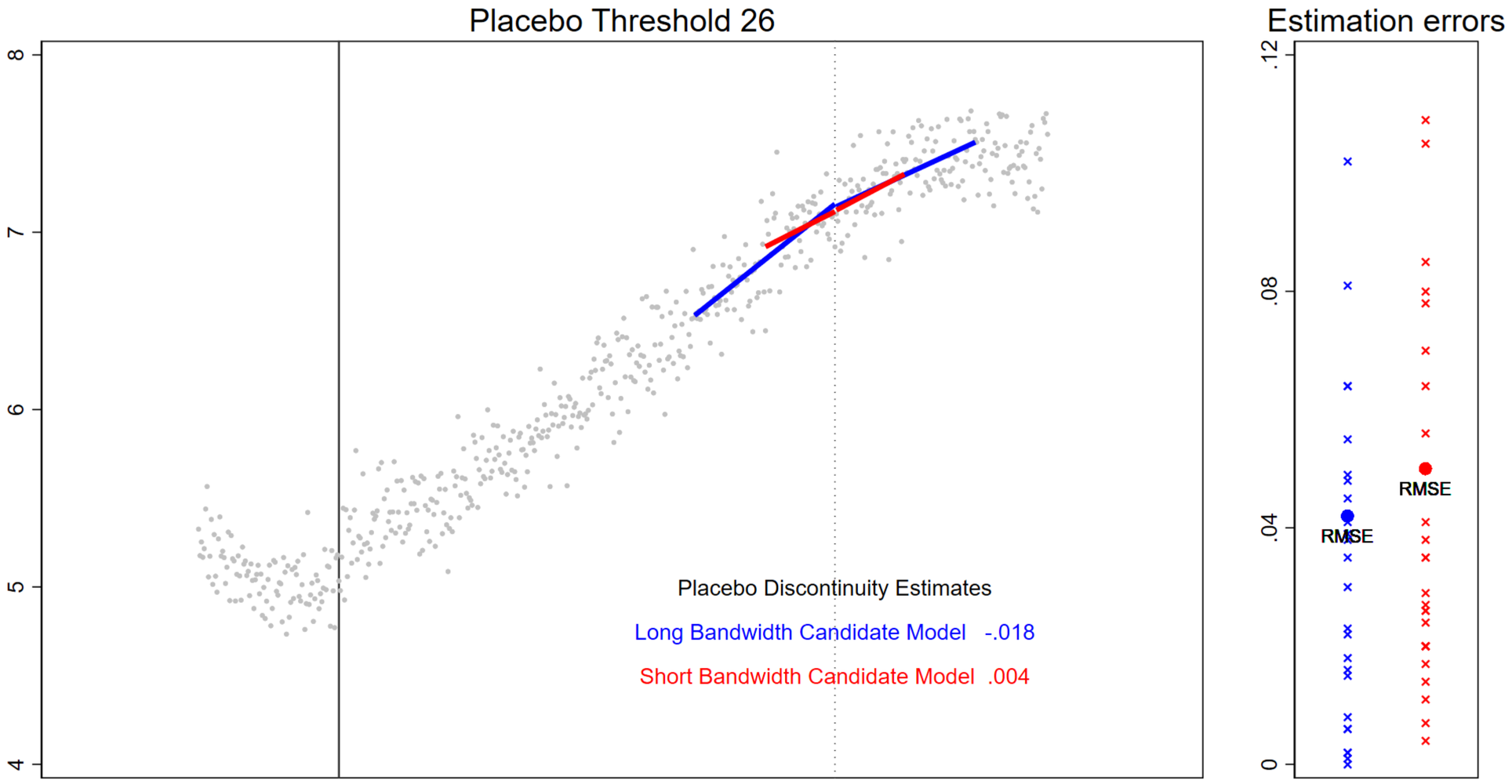

Placebo Threshold 21
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .102
Short Bandwidth Candidate Model .105
Estimation errors
.12
.08
.04
0
RMSE
RMSE

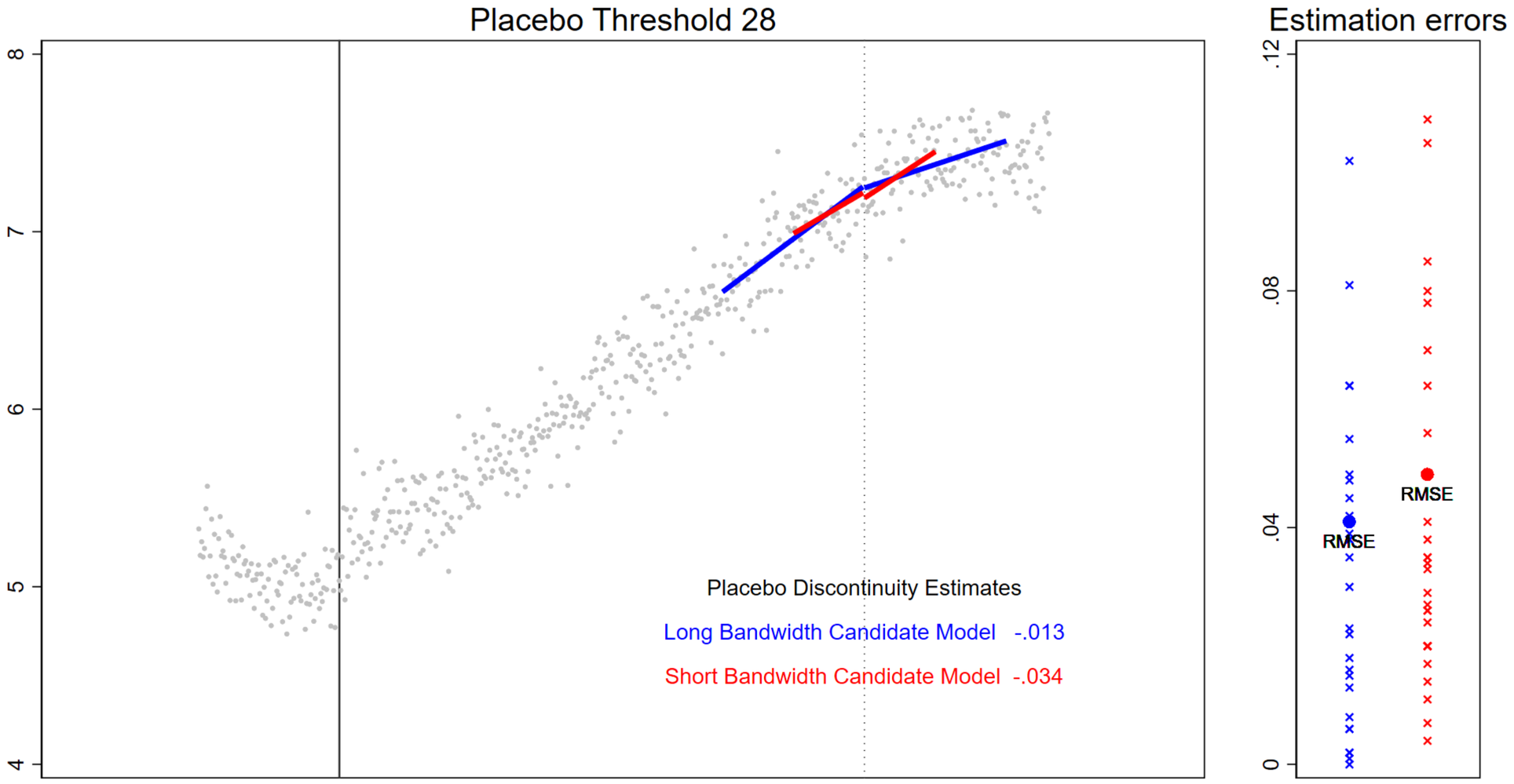

Placebo Threshold 22
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.006
Short Bandwidth Candidate Model -.035
Estimation errors
.12
.08
.04
0
RMSE
RMSE

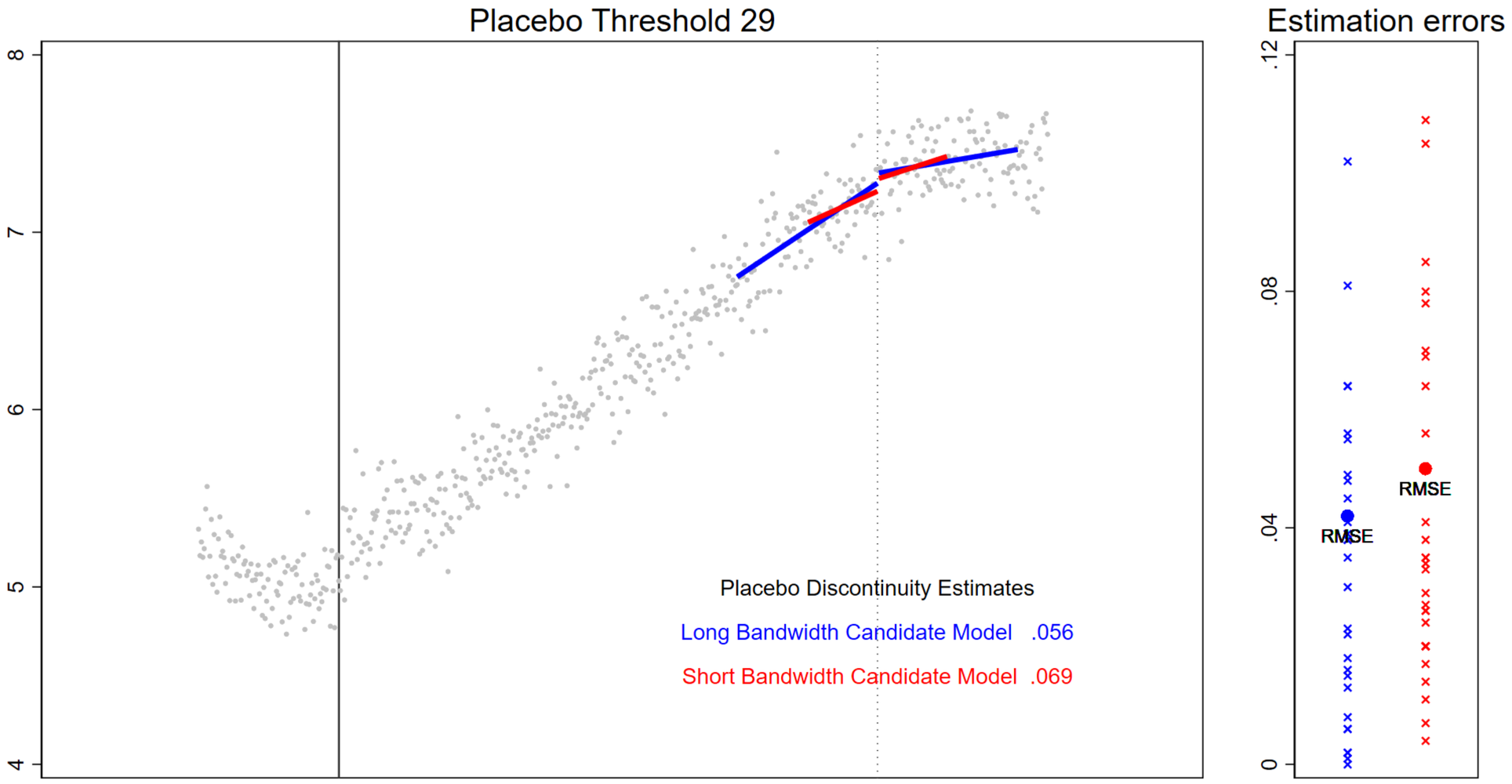

Placebo Threshold 23
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .002
Short Bandwidth Candidate Model -.02
Estimation errors
.12
.08
.04
0
RMSE
RMSE

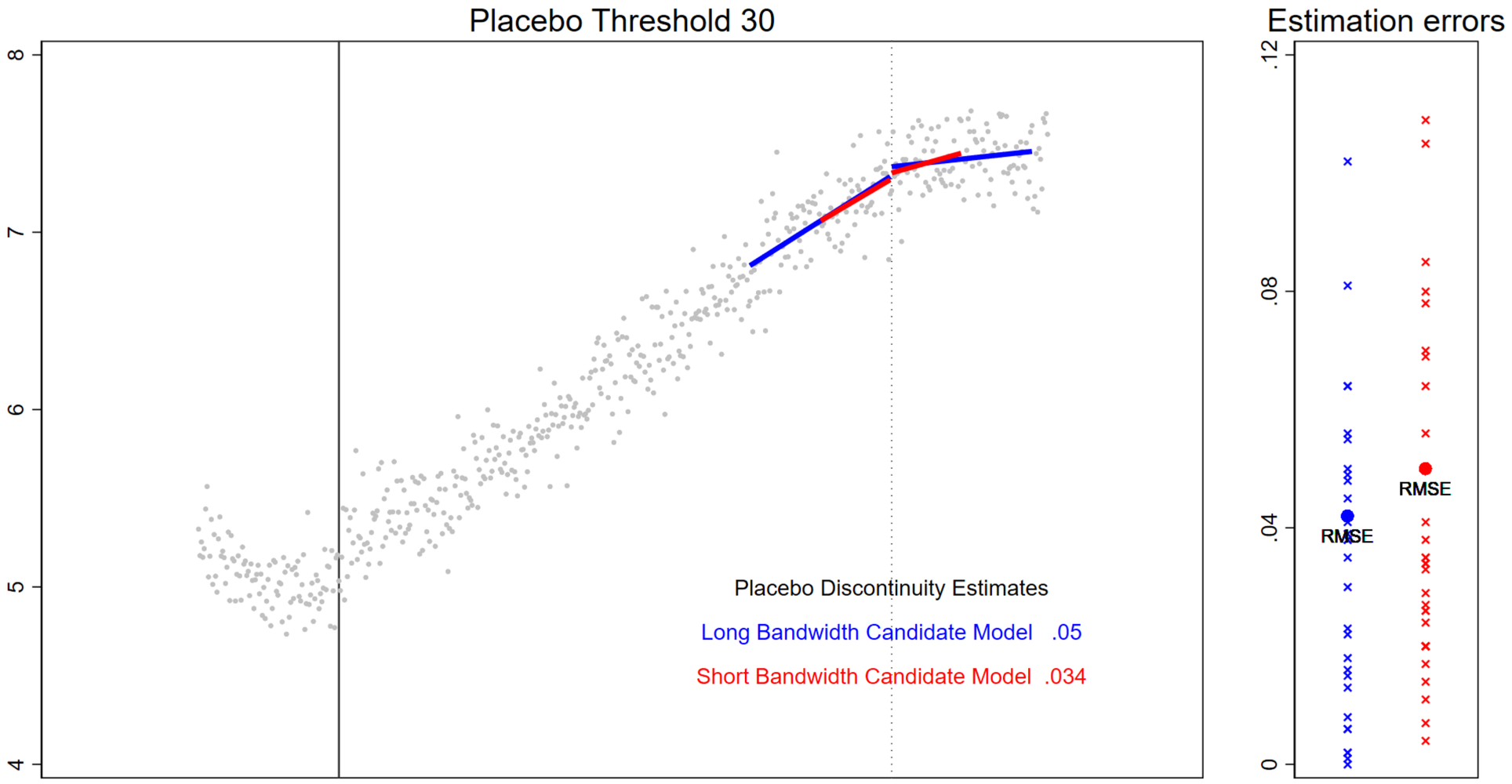

Placebo Threshold 24
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.008
Short Bandwidth Candidate Model -.02
Estimation errors
.12
.08
.04
0
RMSE
RMSE

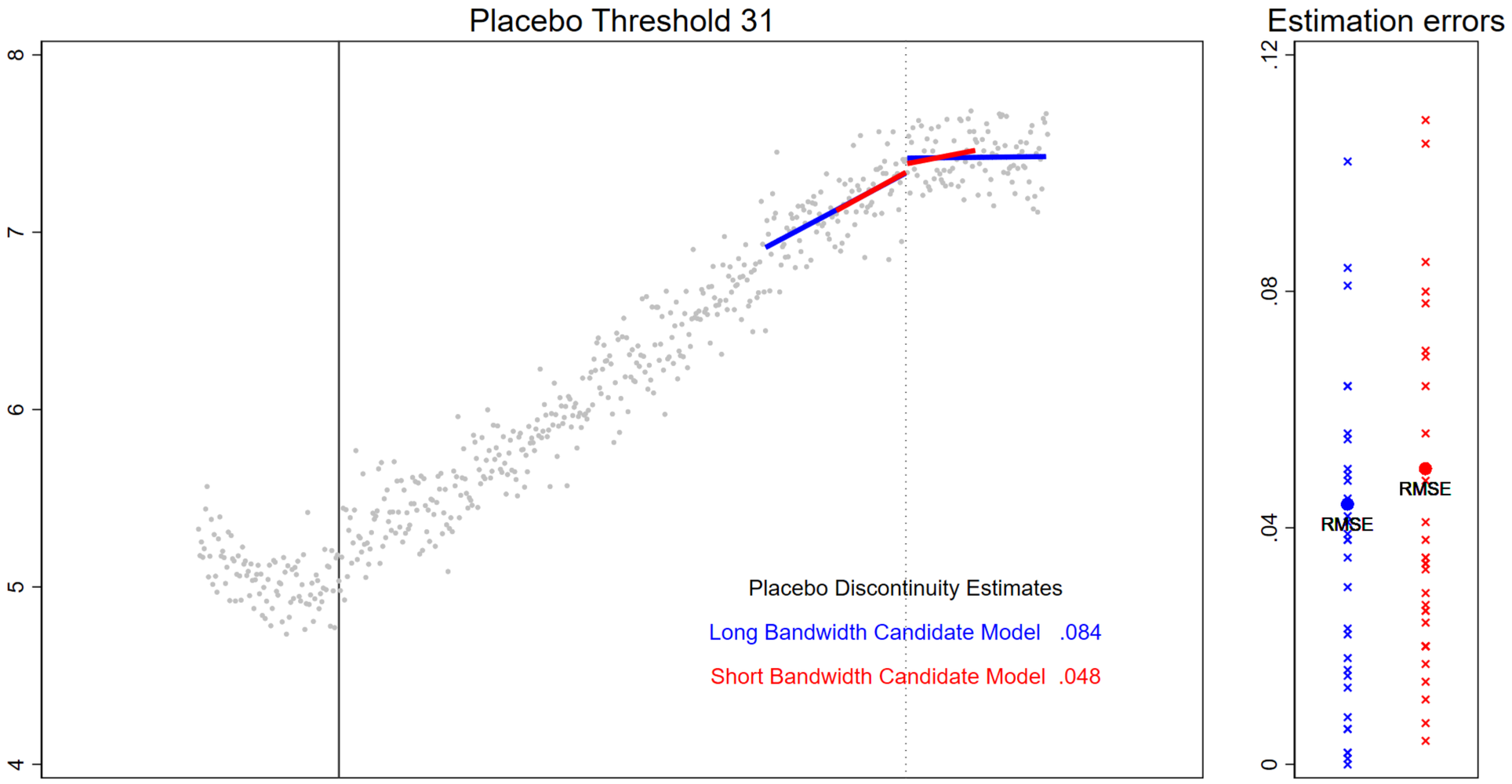

Placebo Threshold 25
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.039
Short Bandwidth Candidate Model -.078
Estimation errors
.12
.08
.04
0
RMSE
RMSE

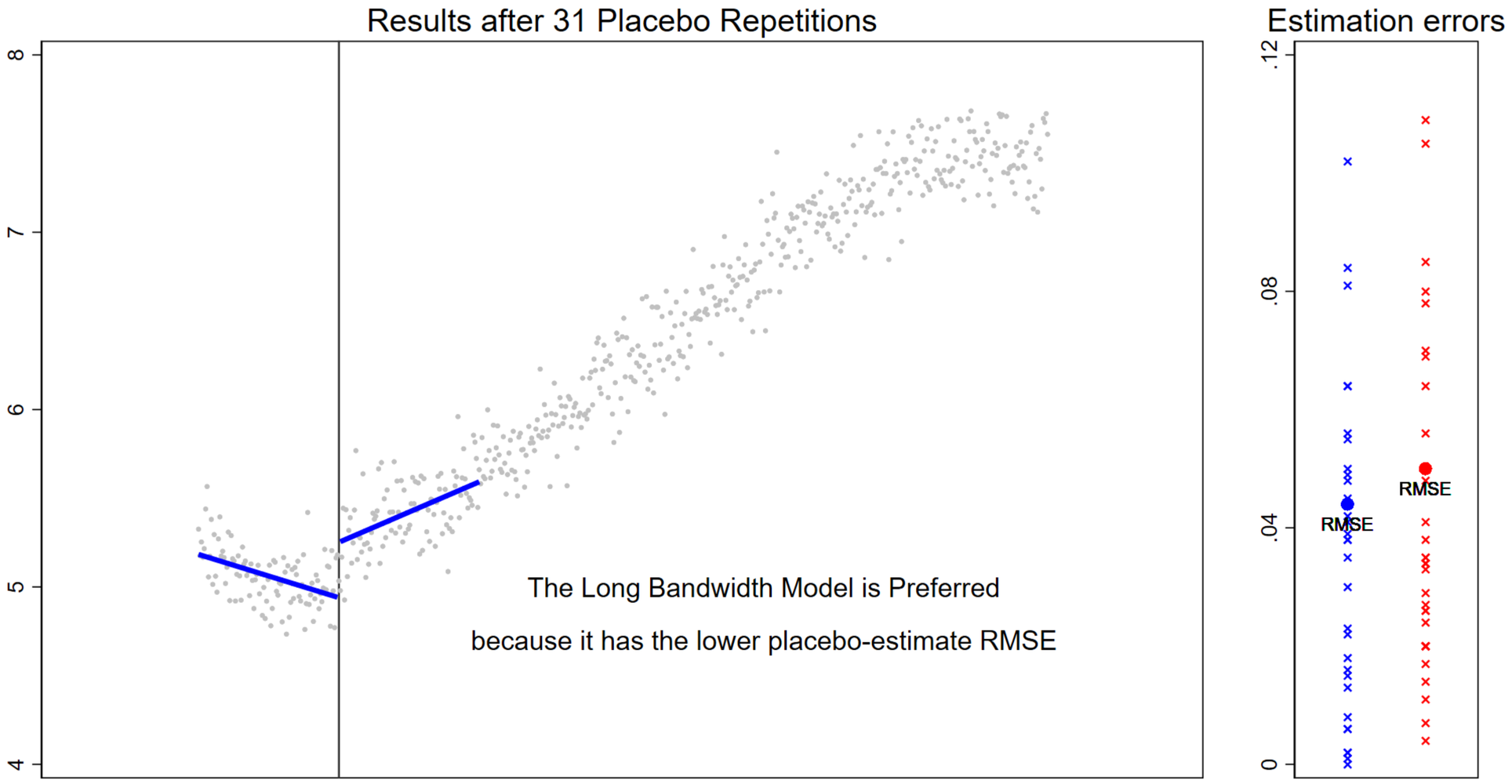

Placebo Threshold 26
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.018
Short Bandwidth Candidate Model .004
Estimation errors
.12
.08
.04
0
RMSE
RMSE

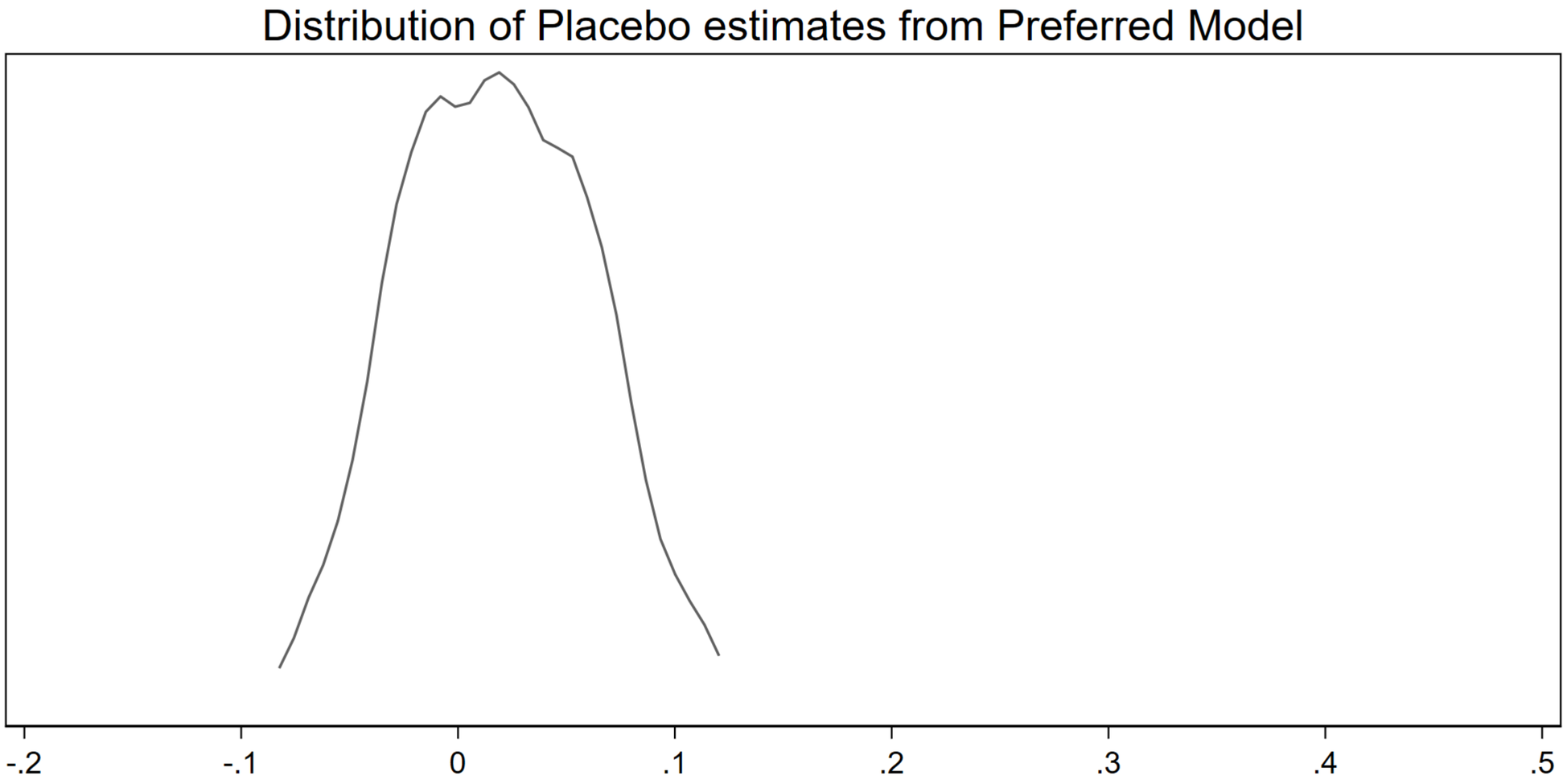

Placebo Threshold 28
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model -.013
Short Bandwidth Candidate Model -.034
Estimation errors
.12
.08
.04
0
RMSE
RMSE

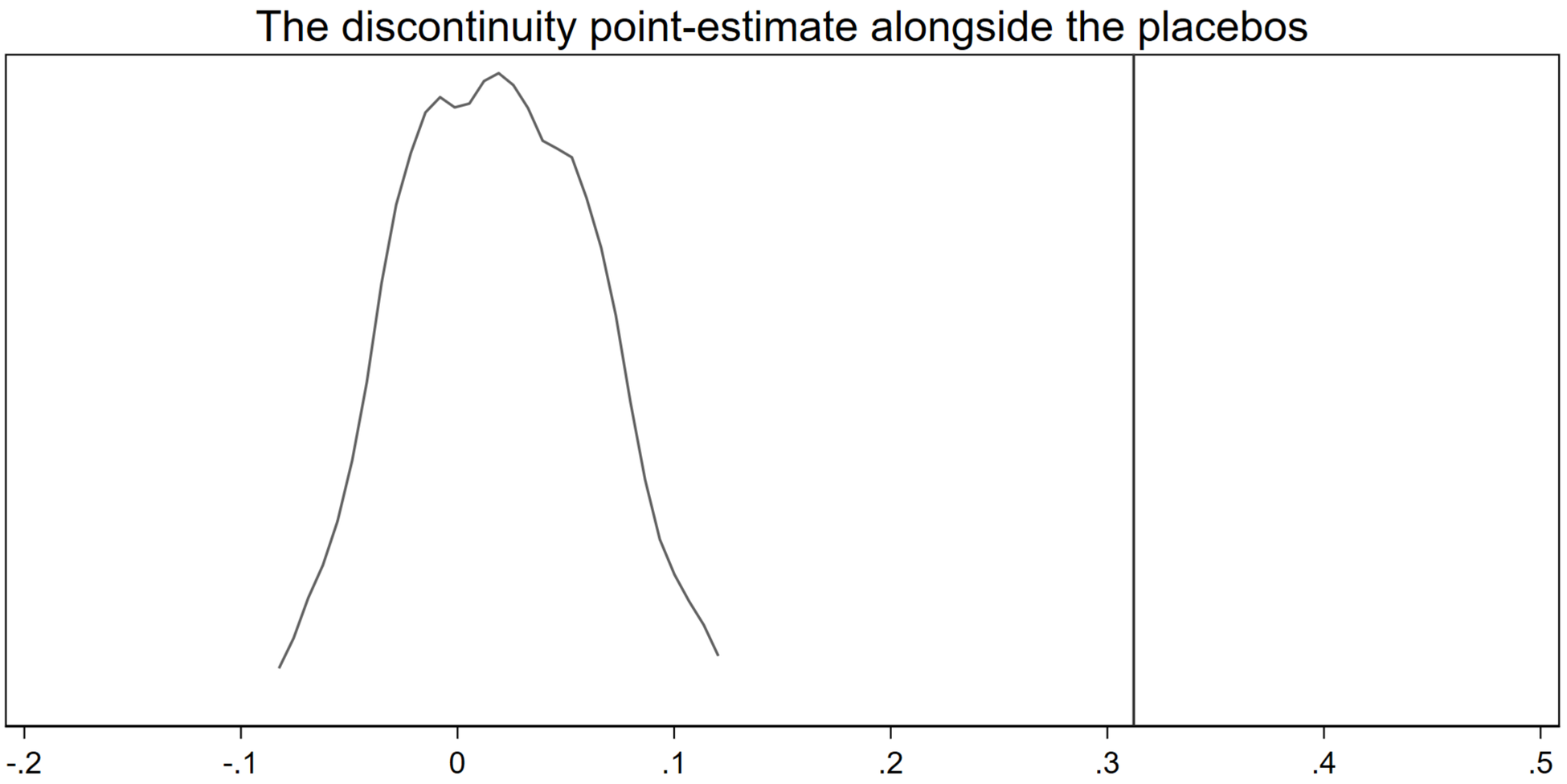

Placebo Threshold 29
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .056
Short Bandwidth Candidate Model .069
Estimation errors
.12
.08
.04
0
RMSE
RMSE

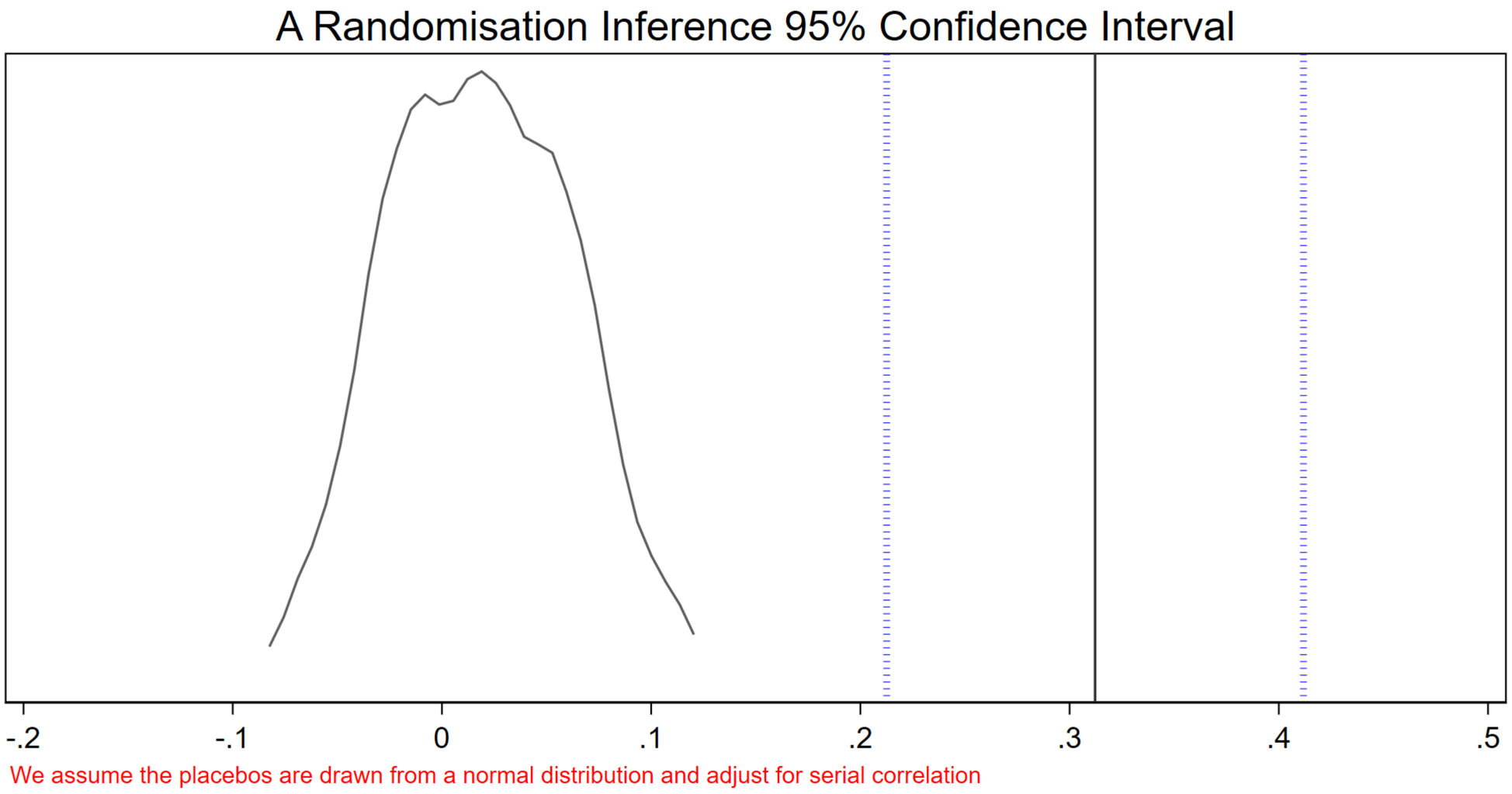

Placebo Threshold 30
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .05
Short Bandwidth Candidate Model .034
Estimation errors
.12
.08
.04
0
RMSE
RMSE


Placebo Threshold 31
8
7
6
5
4
Placebo Discontinuity Estimates
Long Bandwidth Candidate Model .084
Short Bandwidth Candidate Model .048
Estimation errors
.12
.08
.04
0
RMSE
RMSE


Results after 31 Placebo Repetitions
8
7
6
5
4
The Long Bandwidth Model is Preferred
because it has the lower placebo-estimate RMSE
Estimation errors
.12
.08
.04
0
RMSE
RMSE


Distribution of Placebo estimates from Preferred Model
-.2
-.1
0
.1
.2
.3
.4
.5


The discontinuity point-estimate alongside the placebos
-.2
-.1
0
.1
.2
.3
.4
.5



We assume the placebos are drawn from a normal distribution and adjust for serial correlation

Outline

- Key Issues and our Contribution
- **Motivating application – 2 policy changes affecting Learner drivers in NSW**
- Theory – show our approach is asymptotically optimal, under restrictive conditions
- Simulations – our approach performs favourably compared to other procedures using stylised and realistic DGPs

Motivating Example – 2 Policy Changes affecting **Learner Drivers** in NSW

These policy changes were intended to reduce (subsequent) crashes

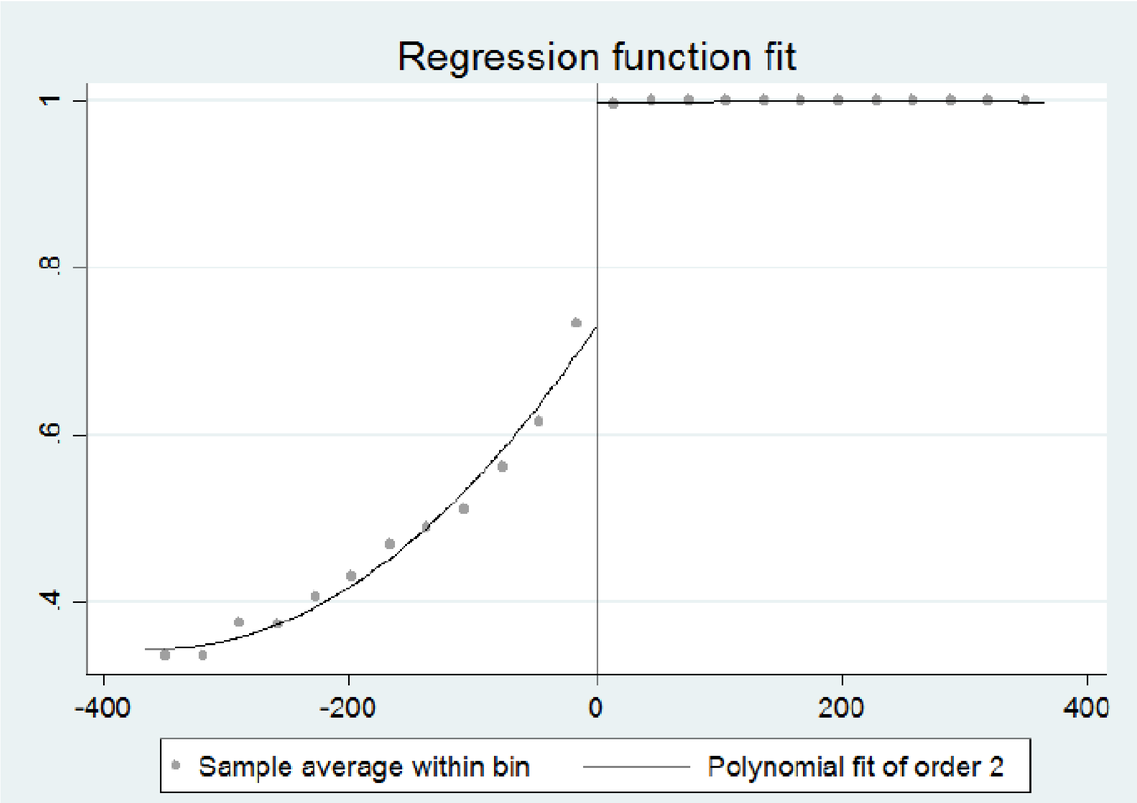
Motivating Example (cont.) These policy changes increased the number of **Mandatory Supervised Driving Hours**, from zero to 50, and then to 120 hours

Motivating Example (cont.)

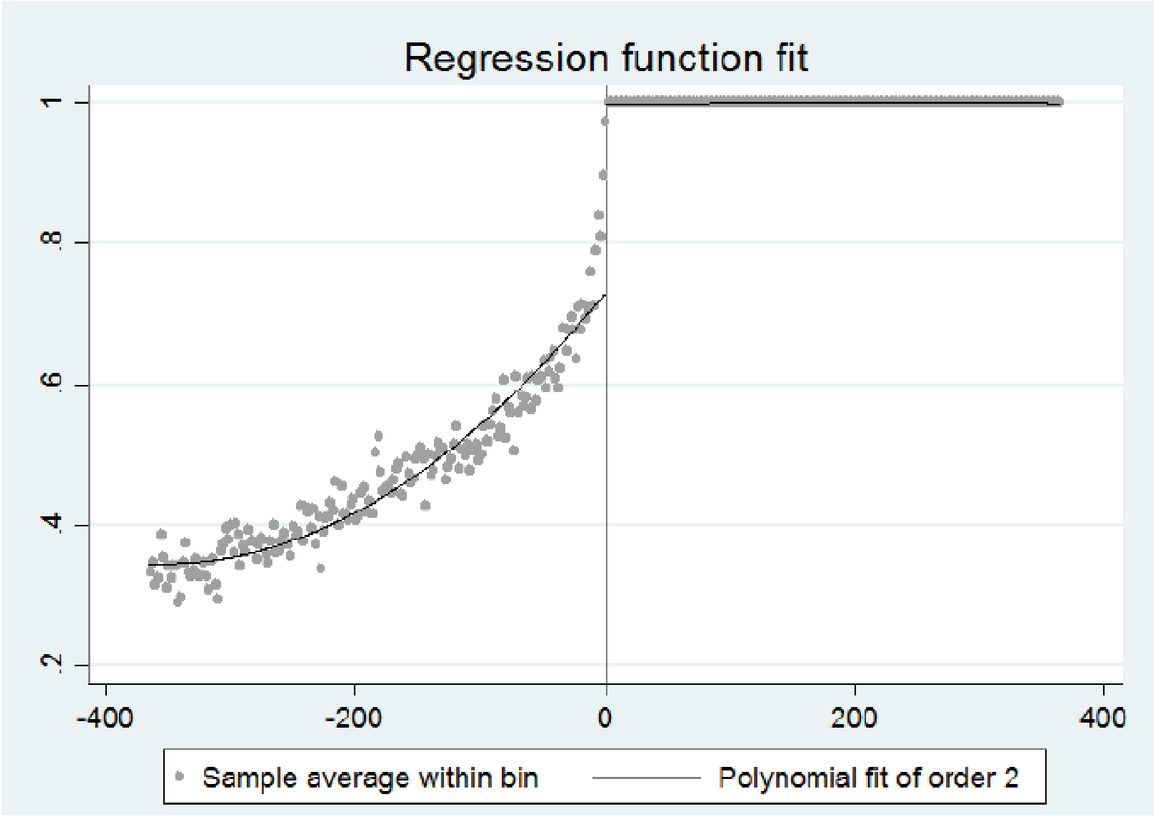
- **Mandatory Supervised Driving Hours (MSDH) for Learner Drivers in NSW**
- On 1 July 2000, MSDH increased from 0 to 50 hours.
- Because of the age requirement for a learner's permit (16 years), people born just before 1 July 1984 could avoid the policy. Those born after 1 July 1984 could not.
- A similar increase occurred on 1 July 2007 from 50 to 120 hours.
- This created discontinuities(?) in the probability of treatment by date of birth (DOB).
- Did these policy changes affect crash rates??

First-stage relationship between DOB (centred around 1 July **1984** and 'Treatment' (i.e. **50+** hours versus 0+)

P(Mandated 50 hours |Ps prior to age 30) - 1 month bins

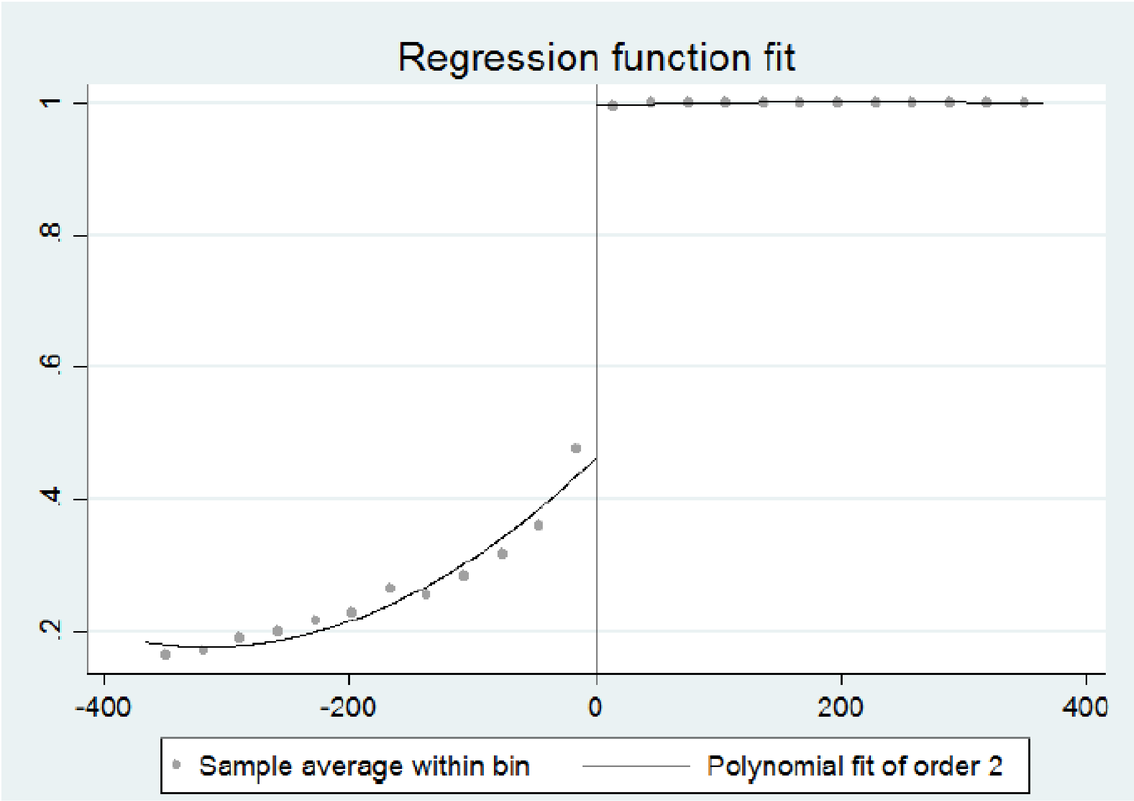


P(Mandated 50 hours |Ps prior to age 30) - 2 day bins

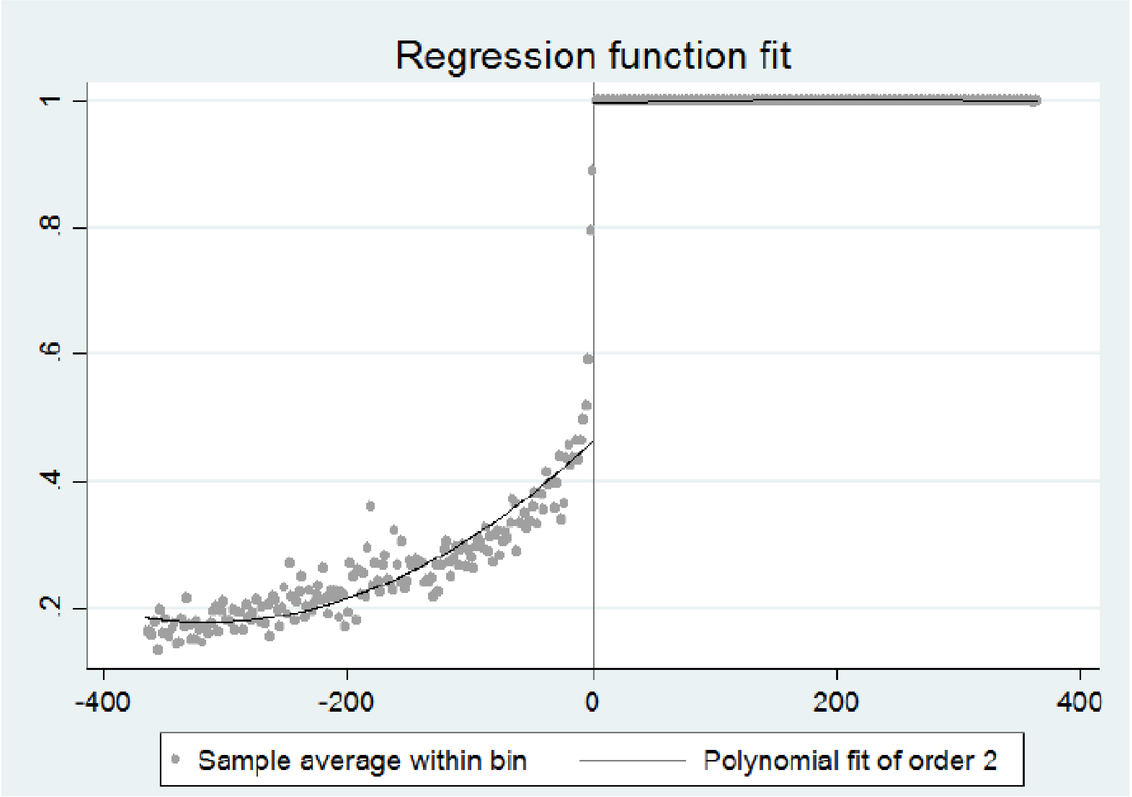


First-stage relationship between DOB (centred around 1 July **1991** and 'Treatment' (i.e. **120+** hours versus 50+)

P(Mandated 120 hours |Ps prior to age 23) - monthly bins



P(Mandated 120 hours |Ps prior to age 23) - monthly bins



Outline

- Key Issues and our Contribution
- Motivating Application – estimating effects of 2 policy changes affecting Learner Drivers in NSW
- **Theory – show our approach is asymptotically optimal, under restrictive conditions**
- Simulations – our approach performs favourably compared to other procedures using stylised and realistic DGPs

Proof of Asymptotic Optimality (Overview)

- Our approach is ‘asymptotically optimal’ if the best treatment effect estimator also has the lowest mean squared placebo estimates, when the placebo zone is large.
- This is the case if the MSE at each placebo threshold equals the MSE at the treatment threshold.
- We show this, assuming the global DGP’s CEF has a zero fourth-derivative, under homoskedasticity, and uniformly distributed x.
- We focus on sharp RDD, with local linear estimators. The results also translate to higher-order polynomials.

Proof of Asymptotic Optimality (Outline)

1. We show that bias of linear RDD estimators is proportional to the 3rd derivative of the DGP's CEF.
2. This bias is constant across the support of the running variable, assuming the 4th derivative is zero. Therefore the bias of placebo estimators equals the bias of the treatment effect estimator
3. The variance of placebo estimators also equals the variance of the treatment effect estimator, assuming homoskedasticity and a uniform distribution
4. For each estimator, the observed mean of squared estimates across the placebo zone approaches the (unobservable) MSE of the treatment effect estimator as the placebo zone becomes large
5. Therefore the approach is 'asymptotically optimal'

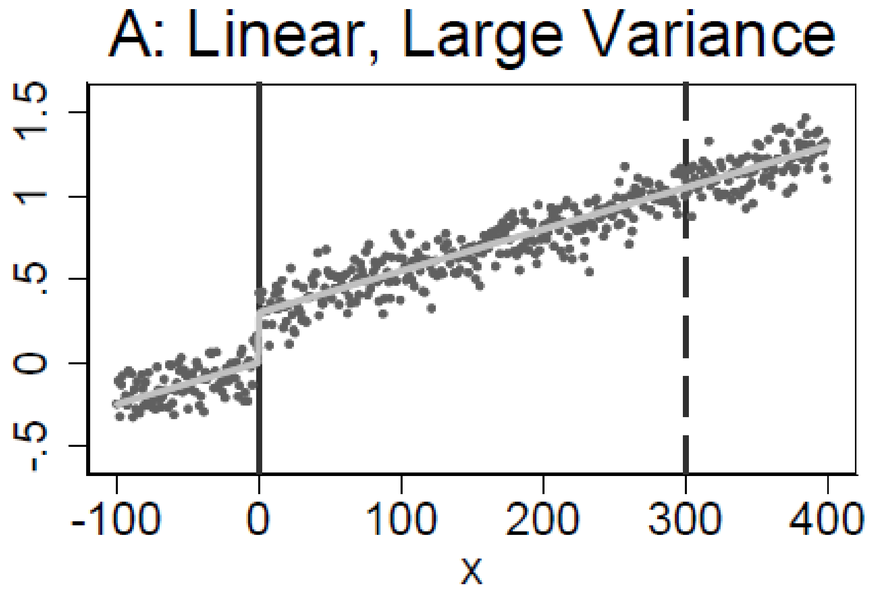
Outline

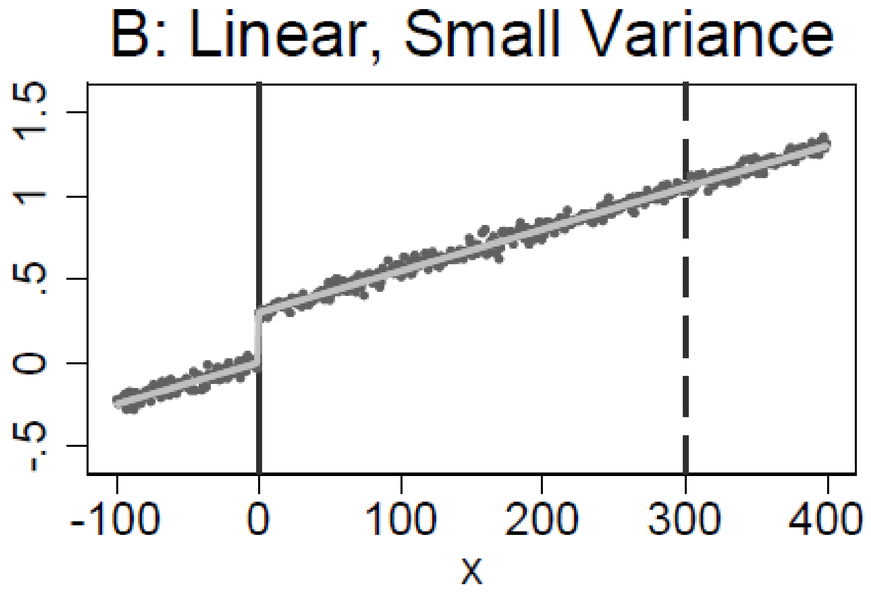
- Key Issues and our Contribution
- Motivating Application – 2 policy changes affecting Learner Drivers in NSW
- Theory – show our approach is asymptotically optimal, under restrictive conditions
- **Simulations – our approach performs favourably compared to other procedures using stylised and realistic DGPs**

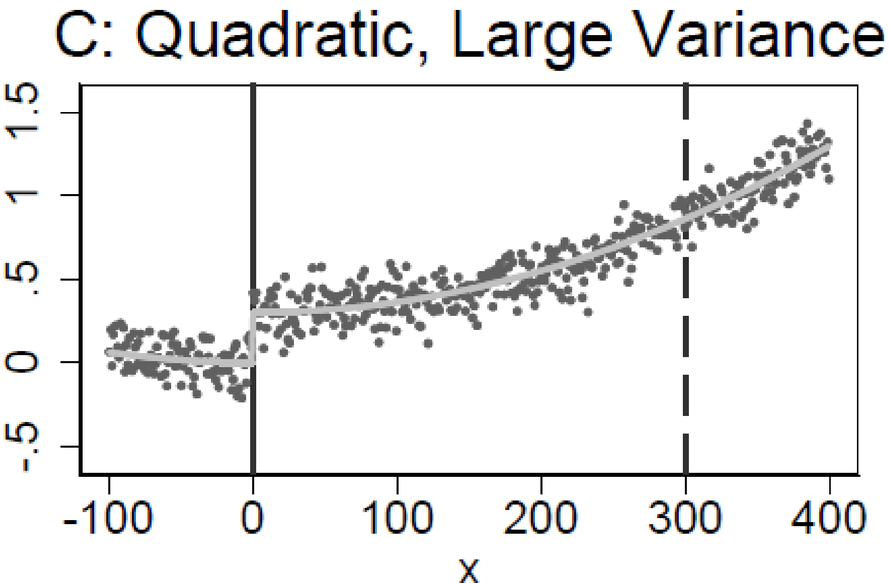
Monte Carlo Simulations

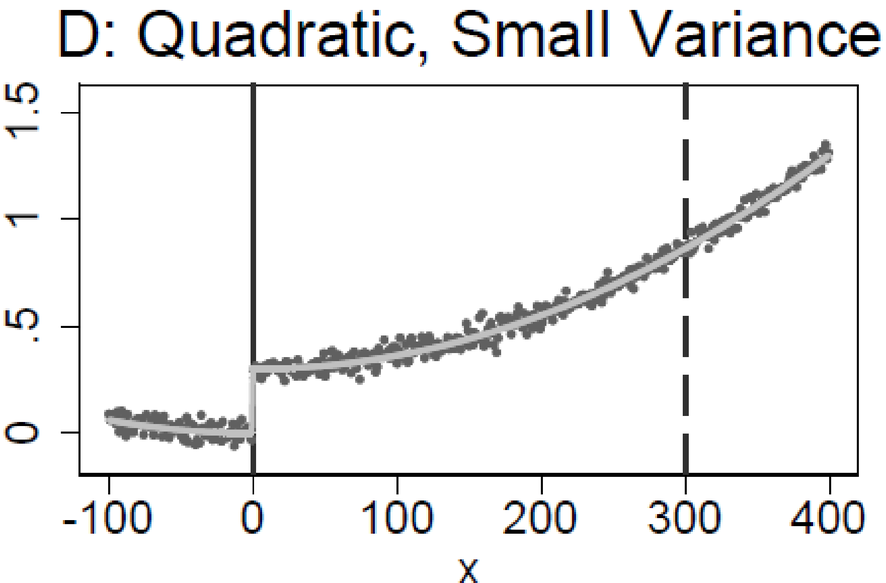
- Great in theory, but does our approach work well with:
  - Finite placebo zones?
  - DGPs with non-zero fourth derivatives?
  - DGPs with non-constant density
  - Realistic DGPs, such as those in prominent well-known RDD studies?
- Our simulation approach closely follows related simulation work
- Use 1000 reps. In each rep, we
  i) trial many candidate estimators (linear and quadratic models with a wide range of BWs) through the placebo zone and pick the best performer on RMSE
  ii) Apply that model to estimate the actual treatment effect
- Compare RMSE (across reps) of our approach to those chosen by the CCT and IK algorithms

Stylised DGPs (first draw of 1000 reps)

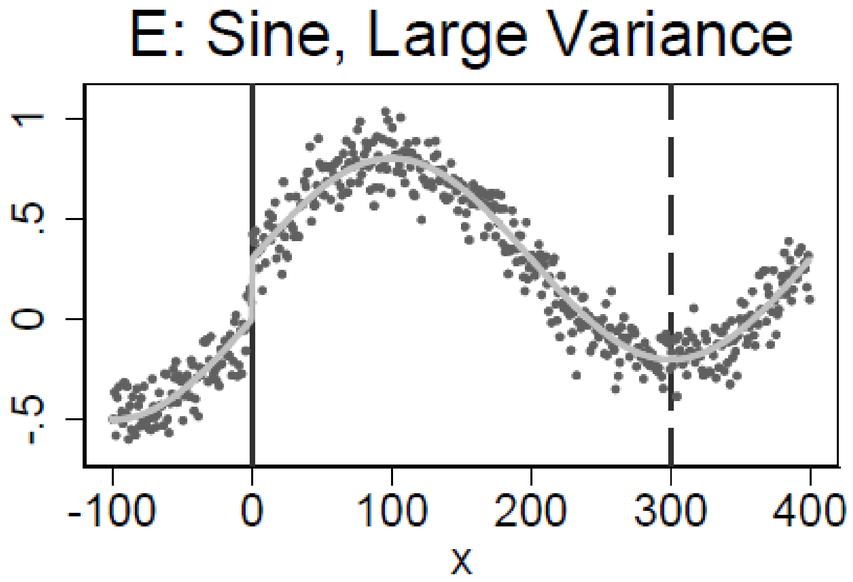


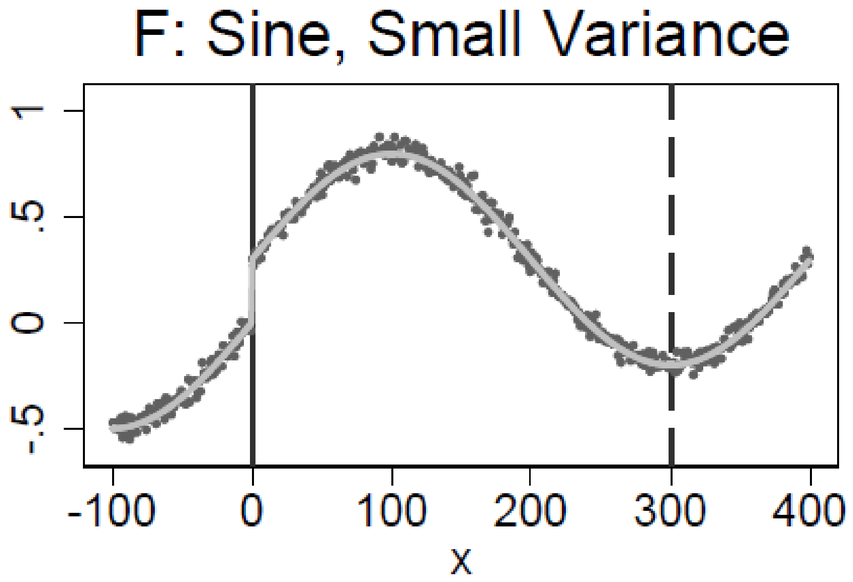


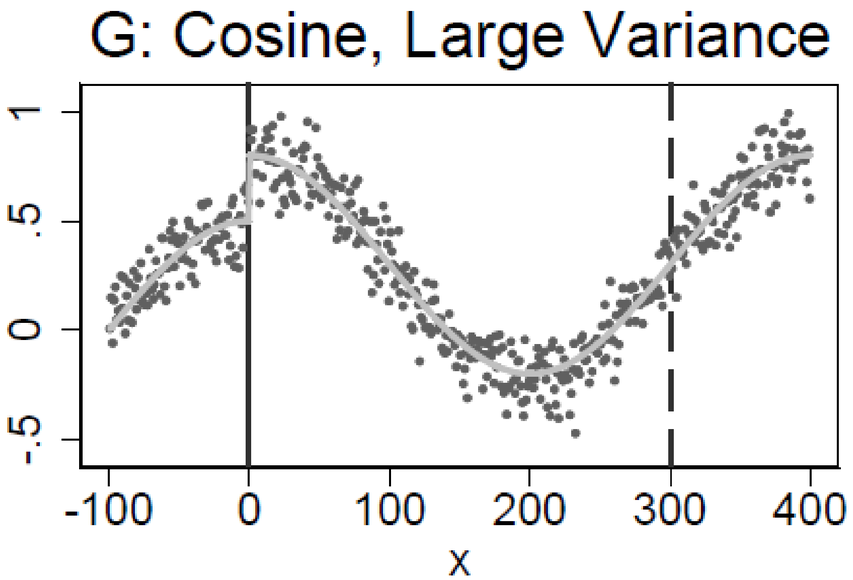


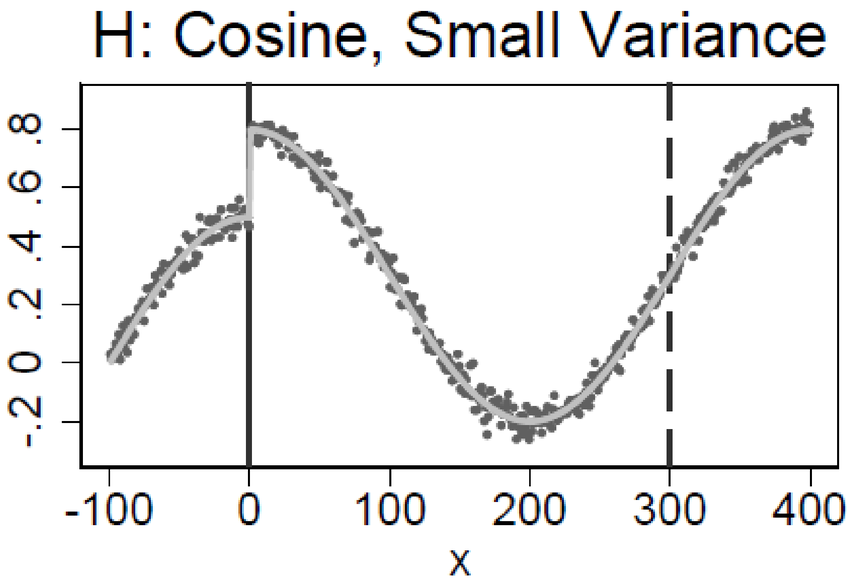


Stylised DGPs (first draw of 1000 reps)







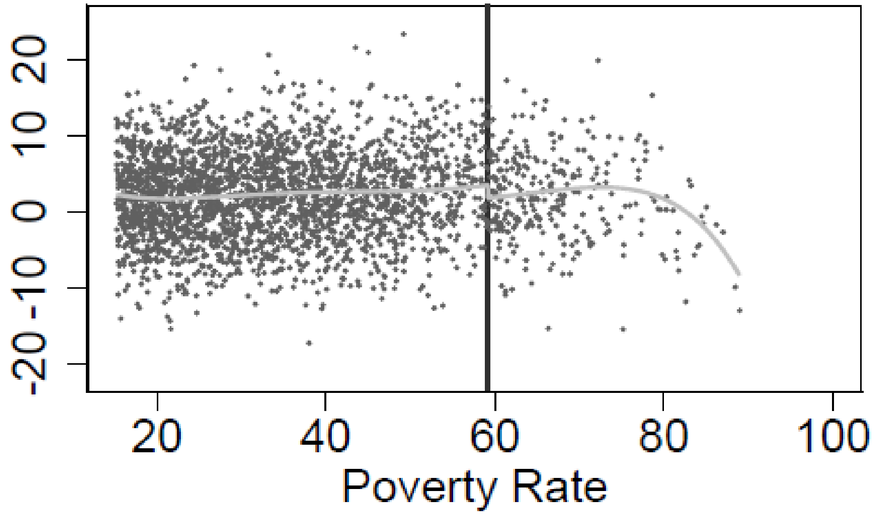


Monte Carlo Simulations (cont.)

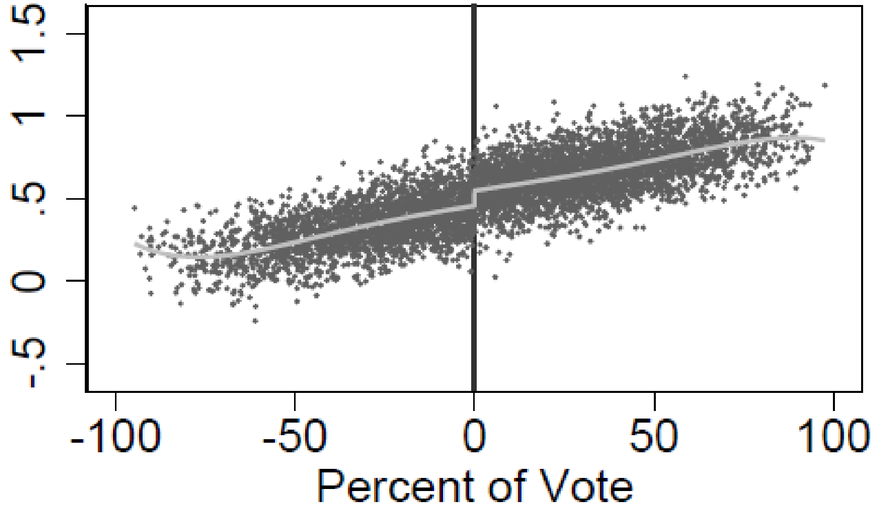
- **Realistic** DGPs
  - Mimic well-known applications: Head Start (Ludwig & Miller, 2007), political incumbency (Lee, 2008), and Minimum Legal Drinking Age (MLDA)
  - Fit f(x), 5^{th} order polynomial to original data, allowing a discontinuity and kink
  - Fit Beta-distribution to summarise distribution of running variable.
  - In each iteration, sample size is set equal to the original sample.
  - Randomly draw values of the running variable from the beta distribution.
  - Set y = f(x) + e, where e is normally distributed with zero mean and variance equal to the variance of the residuals from the regression in 1^{st} step.

Realistic DGPs (first draw of 1000 reps)

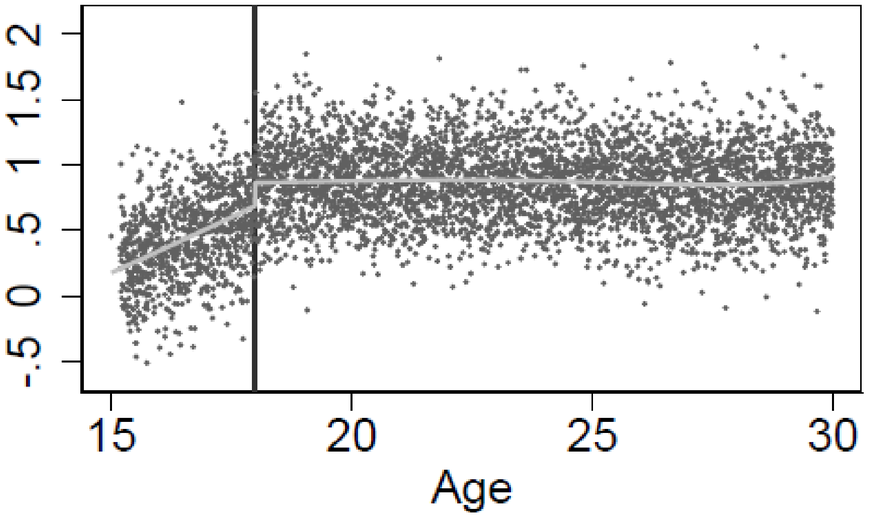
A: Head Start Simulation



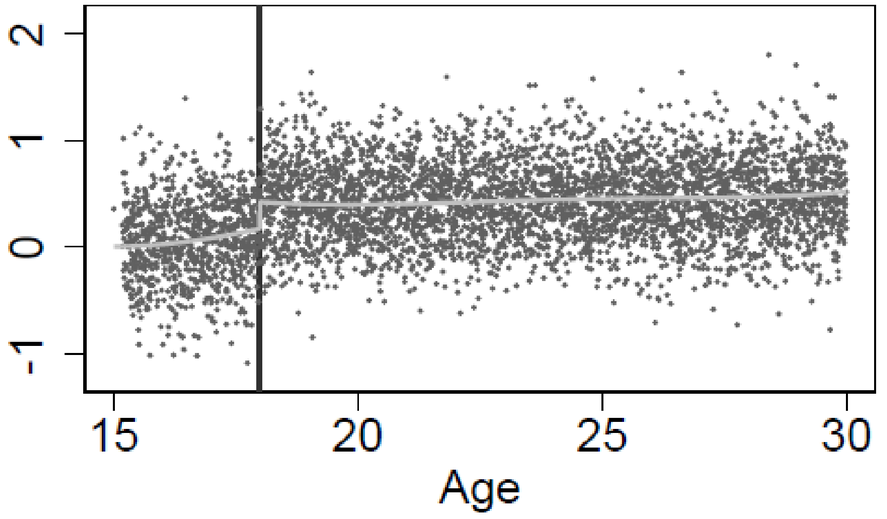
B: Incumbency Simulation



C: MLDA Sim - Ever Drinks



D: MLDA Sim - Drinks Regularly



Monte Carlo Results (4)

	KS			CCT		IK	
	RMSE	mean BW	order (%)	RMSE	mean BW	RMSE	mean BW
			F: Head Start DGP				
Mortality	**0.6794**	**13.86**	**0.999**	1.4265	7.93	0.9853	14.87
			G: Political incumbency DGP				
Wins	**0.0111**	**21.59**	**0.988**	0.0125	22.76	0.0119	30.35
			H: MLDA DGP				
Ever Drinks	**0.0271**	**3.81**	**0.981**	0.0591	0.96	0.0347	2.50
Drinks Regularly	**0.0340**	**3.96**	**0.983**	0.0750	0.96	0.0425	2.89
Proportion of Days Drinks	**0.0154**	**3.96**	**0.983**	0.0339	0.96	0.0189	3.15

Monte Carlo Results (4)

	KS			CCT		IK	
	RMSE	mean BW	order (%)	RMSE	mean BW	RMSE	mean BW
			F: Head Start DGP				
Mortality	**0.6794**	**13.86**	**0.999**	1.4265	7.93	0.9853	14.87
			G: Political incumbency DGP				
Wins	**0.0111**	**21.59**	**0.988**	0.0125	22.76	0.0119	30.35
			H: MLDA DGP				
Ever Drinks	**0.0271**	**3.81**	**0.981**	0.0591	0.96	0.0347	2.50
Drinks Regularly	**0.0340**	**3.96**	**0.983**	0.0750	0.96	0.0425	2.89
Proportion of Days Drinks	**0.0154**	**3.96**	**0.983**	0.0339	0.96	0.0189	3.15

Monte Carlo Results Summary

- Our approach always beats CCT and IK using DGPs based on real data
- Our approach beats CCT in most of the stylized simulations as well, including those with highly unstable DGPs (Sine and Cosine)
  - Performance against IK more mixed (we usually win with simpler DGPs (linear, quadratic), particularly with more error variance, but usually lose with sine/cosine. However, RMSEs and selected BWs similar.

Outline

- Key Issues and our Contribution
- **Motivating application – 2 policy changes affecting Learner drivers in NSW**
- Theory – show our approach is asymptotically optimal, under restrictive conditions
- Simulations – our approach performs favourably compared to other procedures using stylised and realistic DGPs

Candidate models for our application

1. Conventional (fully-interacted) linear RDD.
2. RDD with a linear fit on the right side of the threshold, and a quadratic on the left.
3. Conventional (fully-interacted) quadratic RDD.
4. Conventional (fully-interacted) linear RPJKD.
5. Quadratic RPJKD, in which quadratic term is not interacted with the threshold indicator.
6. RPJKD model with a linear fit on the right side of the threshold, and a quadratic on the left.
7. Fully-interacted quadratic RPJKD.
8. Conventional (fully-interacted) linear RKD.
9. Quadratic RKD, in which the quadratic term is not interacted with the threshold indicator.
10. RKD with a linear fit on the right side of the threshold, and a quadratic on the left.
11. Fully-interacted quadratic RKD.
12. Month-of-birth cohort IV, with linear DOB control
13. Month-of-birth cohort IV, with quadratic DOB control
14. Month-of-birth cohort IV, with cubic DOB control

Summary of Model Performance in Placebo Zone

Model	Description	RMSE	Optimal BW	Coverage	Bias
1	RDD - linear	0.0083	365	0.962	-0.0004
2	RDD - mixed polynomial	0.0199	365	0.921	0.0014
3	RDD - quadratic	0.0230	365	0.927	0.0015
4	**RPJKD - linear**	0.0060	365	0.936	0.0010
5	RPJKD - quadratic	0.0073	365	0.980	-0.0005
6	**RPJKD - mixed polynomial**	0.0052	365	0.992	0.0000
7	RPJKD - interacted quadratic	0.0132	365	0.938	0.0005
8	RKD - linear	0.0096	355	0.910	0.0028
9	RKD – quadratic	0.0179	365	0.953	0.0019
10	**RKD - mixed polynomial**	0.0057	365	0.984	0.0002
11	RKD - interacted quadratic	0.0177	365	0.950	0.0019
12	**birth cohort-IV - linear**	0.0051	365	0.946	0.0006
13	birth cohort-IV - quadratic	0.0070	365	0.987	-0.0007
14	birth cohort-IV - cubic	0.0124	365	0.937	0.0001

Key Estimates

Table 6: Estimated effects of minimum supervised driving hours

	Best estimator	Best sym. cohort-IV	Best sym. RPJKD	Best sym. RKD	Best sym. RDD
	(1)	(2)	(3)	(4)	(5)
	A: 2000 Reform (0 → 50 hours)				
MVA	-0.0144***	-0.0132***	-0.0147***	-0.0144**	-0.0168***
1-year					
SE	0.0041	0.0049	0.0050	0.0058	0.0058
p-value	0.0005	0.0073	0.0032	0.0129	0.0038
alt. SE	0.0039	0.0047	0.0049	0.0054	0.0080
alt. p-value	0.0096	0.0203	0.0157	0.0381	0.0552
Model	6	12	6	10	1
BW	365 / 550	365	365	365	365
	B: 2007 Reform (50 → 120 hours)				
MVA	0.0021	0.0003	0.0006	-0.0024	-0.0007
1-year					
SE	0.0030	0.0033	0.0033	0.0046	0.0035
p-value	0.4790	0.9259	0.8477	0.6069	0.8422
alt. SE	0.0042	0.0047	0.0049	0.0054	0.0080
alt. p-value	0.6310	0.9496	0.9003	0.6774	0.9328
Model	12	12	6	10	1
BW	560 / 365	365	365	365	365

Conclusions

- We propose a new approach for model selection in RDD and related designs using placebo zone data
- Can compare across and within model types, on any number of dimensions. Also offers a new approach for inference
- We recommend its use whenever the DGP is `stable' across the range of the running variable, and the placebo zone is not small

Policy conclusions

- Going from 0 → 50 MSDH reduced MVAs. Relatively large effect sizes (21% reduction in 1st year). No effect after 1-2 years (evidence against habit formation).
- No effect from further increase to 120 MSDH.

Our 2 Stata Programs

ssc install pzms

https://econpapers.repec.org/software/bocbocode/s459073.htm

-pzms- implements our approach. Very easy to use. Only required option is the maximum bandwidth

pzms_sim uses simulations based on the data from any application, to examine likely performance of our approach, compared to alternative approaches https://sites.google.com/site/nrkettlewell/research

Thank you

Extra Slides

Data

- Administrative records on licence history of drivers in NSW linked with records on crashes between 1996 and 2016. Only crashes where at least one car is towed away and/or someone was injured are recorded.
- Because of important discontinuous policy changes we only consider BWs of up to 365 days for our baseline analysis.
- In main analysis N = 154,524 drivers born within 1 year of 1 July 1984 (2000 reform), N = 160,301 within 1 year of 1 July 1991 (2007 reform).

What are Researchers Doing?

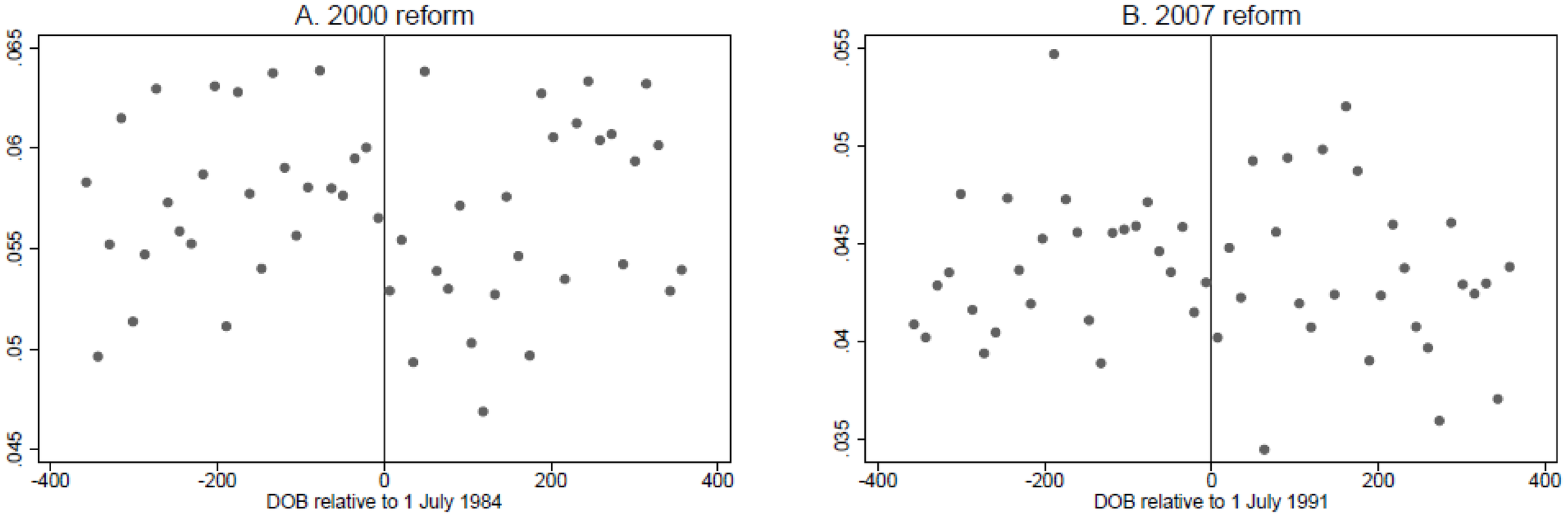
Table: Discontinuity studies published in leading journals in 2019

	Sharp RDD	Fuzzy RDD	Cohort-IV
Papers using this model	15	10	2
Method for bandwidth choice			
No stated method	6	4	2
IK/CCT	8	6	0
Method for polynomial choice			
No stated method	10	9	2
Local linear polynomial as baseline	11	8	-
Robustness tests			
Varied bandwidth	15	9	1
Varied polynomial	13	5	1

Our Approach

- Estimate many placebo treatment effects for each candidate estimator by moving the placebo threshold across the placebo zone where we know the treatment effect = 0.
- Choose the estimator with lowest root mean squared error (RMSE) of treatment effect estimates across all candidate models. Can also assess bias and coverage.
- Can compare models on any dimension and of different model types.
- Extendable to nonlinear models (e.g. logit)
- Combine with randomization inference similar to Ganong & Jager (2018)

Figure: Reduced-Form Relationships between DOB and MVA 1-year



Notes: Scatter plots use 14-day bin size.

Decisions to make

- Which empirical technique should we use (for estimation)?
- Which order of polynomial should we use?
- What bandwidth should we use?
- Should we use a different approach on each side of the threshold?

Monte Carlo Simulations (cont.)

- **Stylised** DGPs
  - Sample size = 500 observations (following IK, CCT and others)
  - x is uniformly distributed across the range (-100, 400)
  - Outcome variable y = 0:3(x>0) + f(x) + e,
  - where 0.3 is the discontinuity at x = 0, and $\sigma^2 = 0.1^2$ (representing `large' error variance), or $0.3^2$ (`small' error variance)
  - f(x) is either: Linear, Quadratic, Sine, or Cosine

Proof of Asymptotic Optimality (1)

Let τ be the true treatment effect. The MSE of any estimator is

$$MSE(\hat{\tau}) = E(\hat{\tau} - \tau)^2 = Var(\hat{\tau}) + Bias(\hat{\tau})^2 \quad (1)$$

Consider local linear RD estimators with bandwidth b. Assume n_b observations within this bandwidth, on each side of the threshold.

$\hat{\tau}_b = \hat{\alpha}_{2b} - \hat{\alpha}_{1b}$, where $\hat{\alpha}_{2b}$ and $\hat{\alpha}_{1b}$ are estimated using independent linear regressions on each side of the threshold ($x = 0$)

$$y = \alpha_1 + \beta_1 x + \varepsilon, \qquad -b < x < 0 \quad (2a)$$

$$y = \alpha_2 + \beta_2 x + \varepsilon, \qquad 0 < x < b \quad (2b)$$

$$MSE(\hat{\tau}_b) = E(\hat{\alpha}_{2b} - \hat{\alpha}_{1b} - \tau)^2 = Var(\hat{\alpha}_{1b}) + Var(\hat{\alpha}_{2b}) + Bias(\hat{\tau}_b)^2 \quad (3)$$

Proof of Asymptotic Optimality (2)

Consider a true DGP that is cubic: $y = \alpha + \tau I(x > 0) + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \varepsilon,$ (4)

$\theta_2 x^2$ and $\theta_3 x^3$ are omitted variables from (2a) and (2b), $\alpha_1 = \alpha$ and $\alpha_2 = \alpha + \tau$. Using conventional OVB formulas:

$$E(\hat{\alpha}_{1b}) = \alpha + \hat{\delta}_{L1} + \hat{\delta}_{L2} \quad (5)$$

Where $\hat{\delta}_{L1}$ is the estimated constant in the regression of $\theta_2 x^2$ on x: $\theta_2 x^2 = \delta_1 + \pi_1 x + \epsilon_1, -b < x < 0$ (6A)

And $\hat{\delta}_{L2}$ is the estimated constant in the regression of $\theta_3 x^3$ on x, $\theta_3 x^3 = \delta_2 + \pi_2 x + \epsilon_2, \ -b < x < 0$ (6B)

Similarly for $E(\hat{\alpha}_{2b})$ using data on the RHS:

$$E(\hat{\alpha}_{2b}) = (\alpha + \tau) + \hat{\delta}_{R1} + \hat{\delta}_{R2} \quad (5B)$$

Proof of Asymptotic Optimality (3)

The expected value of the RD estimate is hence:

$$E(\hat{\tau}_b) = E(\hat{\alpha}_{2b}) - E(\hat{\alpha}_{1b}) = (\alpha + \tau) + \hat{\delta}_{R1} + \hat{\delta}_{R2} - (\alpha + \hat{\delta}_{L1} + \hat{\delta}_{L2}) \quad (7)$$

However, $\hat{\delta}_{R1} = \hat{\delta}_{L1}$ and $\hat{\delta}_{R2} = -\hat{\delta}_{L2}$, and so:

$$E(\hat{\tau}) = \tau + 2\hat{\delta}_{L2}$$

(to see this, replace x with $-x$ in (6A) and (6B), noting the assumed uniform distribution of x. (6A) becomes $\theta_2 x^2 = \delta_1 - \pi_1 x + \epsilon_1$. This regression yields exactly the same estimate of δ_1. (6B) becomes $-\theta_3 x^3 = \delta_2 - \pi_2 x + \epsilon_2$. This regression yields an estimated constant exactly equal to $-\hat{\delta}_{L2}$.)

(6B) implies that $\hat{\delta}_{L2}$ is proportional to θ_3, and unrelated to any other parameters of the cubic DGP. The bias of $\hat{\tau}_b$ is hence proportional to the third derivative of the DGP CEF.

Proof of Asymptotic Optimality (4)

The discontinuity estimates at any placebo threshold at $x = k$, have the same bias, assuming the same global DGP, for any $|k| > b$

The linear RD estimate with bandwidth b at a placebo discontinuity at $x = k$ is $\hat{\tau}_{kb} = \hat{\alpha}_{k2b} - \hat{\alpha}_{k1b}$, where $\hat{\alpha}_{k1b}$ and $\hat{\alpha}_{k2b}$ are the estimates from these regressions:

$y = \alpha_{k1} + \beta_1(x - k) + \varepsilon, \quad (k - b) < x < k$ and $y = \alpha_{k2} + \beta_2(x - k) + \varepsilon, \qquad k < x < (k + b)$

Substituting $x_k = x - k$, these regressions are equivalent to

$y = \alpha_{k1} + \beta_1 x_k + \varepsilon, \quad -b < x_k < 0$ (9a), and

$y = \alpha_{k2} + \beta_2 x_k + \varepsilon, \qquad 0 < x_k < b$ (9b)

The DGP can be expressed as $y = \alpha + \theta_1(x_k + k) + \theta_2(x_k + k)^2 + \theta_3(x_k + k)^3 + \varepsilon$ (10)
if $k > b$, and similarly if $k < -b$

Proof of Asymptotic Optimality (5)

Equivalently, $y = \pi_0 + \pi_1 x_k + \pi_2 {x_k}^2 + \theta_3 {x_k}^3$, (11)

Where $\pi_0 = \alpha + k\theta_1 + k^2\theta_2 + k^3\theta_3$, $\pi_1 = \theta_1 + 2k\theta_2 + 3k^2\theta_3$, $\pi_2 = \theta_2 + 3k\theta_3$

(9a), (9b) and (11) are equivalent to equations (2a), (2b) and (4), respectively, with $\tau = 0$ and the threshold at $x_k = 0$. As shown, the bias of the RDD estimate is proportional only to the third derivative of the true DGP's CEF. The third derivative ($6\theta_3$) is the same in (11) as in (4), and so $Bias(\hat{\tau}_{kb}) = Bias(\hat{\tau}_b)$, for $|k| > b$.

It is trivial to show that $Var(\hat{\tau}_{kb}) = Var(\hat{\tau}_b)$. Therefore $MSE(\hat{\tau}_{kb}) = MSE(\hat{\tau}_b)$.

Proof of Asymptotic Optimality (6)

For any given estimator, as the placebo zone gets large, the mean of the squared placebo estimates approaches the MSE of the treatment effect estimator:

Since $\tau_{kb} = 0$, $\lim_{m\to\infty} \frac{1}{m} \sum_{k=b+1}^{b+m} \hat{\tau}_{kb}{}^2 = MSE(\hat{\tau}_b)$

If the DGP has a non-zero fourth derivative, our approach is no longer asymptotically optimal. To see this, assume a fourth-order polynomial DGP and follow the same steps. The equivalent of equation (11) would have a different coefficient of ${x_k}^3$ for each k. Therefore the bias of placebo estimates would be different at each placebo threshold.

Monte Carlo Results (1)

	KS			CCT		IK	
	RMSE	mean BW	linear (%)	RMSE	mean BW	RMSE	mean BW
			A: Linear DGP				
Baseline DGP	**0.0241**	**250.19**	**1.000**	0.0413	66.42	0.0295	149.85
Small error variance	**0.0072**	**250.19**	**1.000**	0.0092	83.01	0.0079	201.92
Small placebo zone	**0.0292**	**151.53**	**0.974**	0.0601	34.70	0.0319	120.94
Small placebo zone and error variance	0.0088	151.53	0.974	0.0180	34.69	**0.0087**	**149.19**
			B: Quadratic DGP				
Baseline DGP	**0.0285**	**114.30**	**0.999**	0.0413	66.37	0.0331	137.67
Small error variance	**0.0088**	**98.63**	**0.995**	0.0124	65.87	0.0166	149.54
Small placebo zone	**0.0318**	**118.92**	**0.969**	0.0601	34.70	0.0332	115.63
Small placebo zone and error variance	**0.0096**	**98.54**	**0.953**	0.0180	34.70	0.0116	117.95

Monte Carlo Results (2)

	KS			CCT		IK	
	RMSE	mean BW	order (%)	RMSE	mean BW	RMSE	mean BW
	C: Cubic DGP						
Baseline DGP	**0.0301**	**93.59**	**1.000**	0.0414	66.38	0.0325	114.81
Small error variance	**0.0104**	**79.49**	**1.000**	0.0125	65.94	0.0109	89.25
Small placebo zone	**0.0321**	**100.39**	**0.953**	0.0601	34.69	0.0340	106.87
Small placebo zone and error variance	**0.0110**	**82.66**	**0.919**	0.0180	34.66	0.0111	83.62
	D: Sine DGP						
Baseline DGP	0.0481	59.0	0.857	0.0605	34.1	**0.0443**	**60.8**
Small error variance	**0.0172**	**41.6**	**0.886**	0.0187	30.8	0.0174	56.5
Small placebo zone	0.0633	79.9	0.729	**0.0605**	**34.1**	**0.0443**	**60.8**
Small placebo zone and error variance	0.0207	58.6	0.673	**0.0187**	**30.8**	**0.0174**	**56.5**

Monte Carlo Results (3)

	KS			CCT		IK	
	RMSE	mean BW	order (%)	RMSE	mean BW	RMSE	mean BW
	E: Cosine DGP						
Baseline DGP	0.0446	83.9	0.640	0.0601	34.7	**0.0377**	**75.1**
Small error variance	0.0147	42.1	0.876	0.0180	34.7	**0.0114**	**71.0**
Small placebo zone	0.0437	82.5	0.579	0.0601	34.7	**0.0377**	**75.1**
Small placebo zone and error variance	0.0153	41.8	0.778	0.0180	34.7	**0.0114**	**71.0**

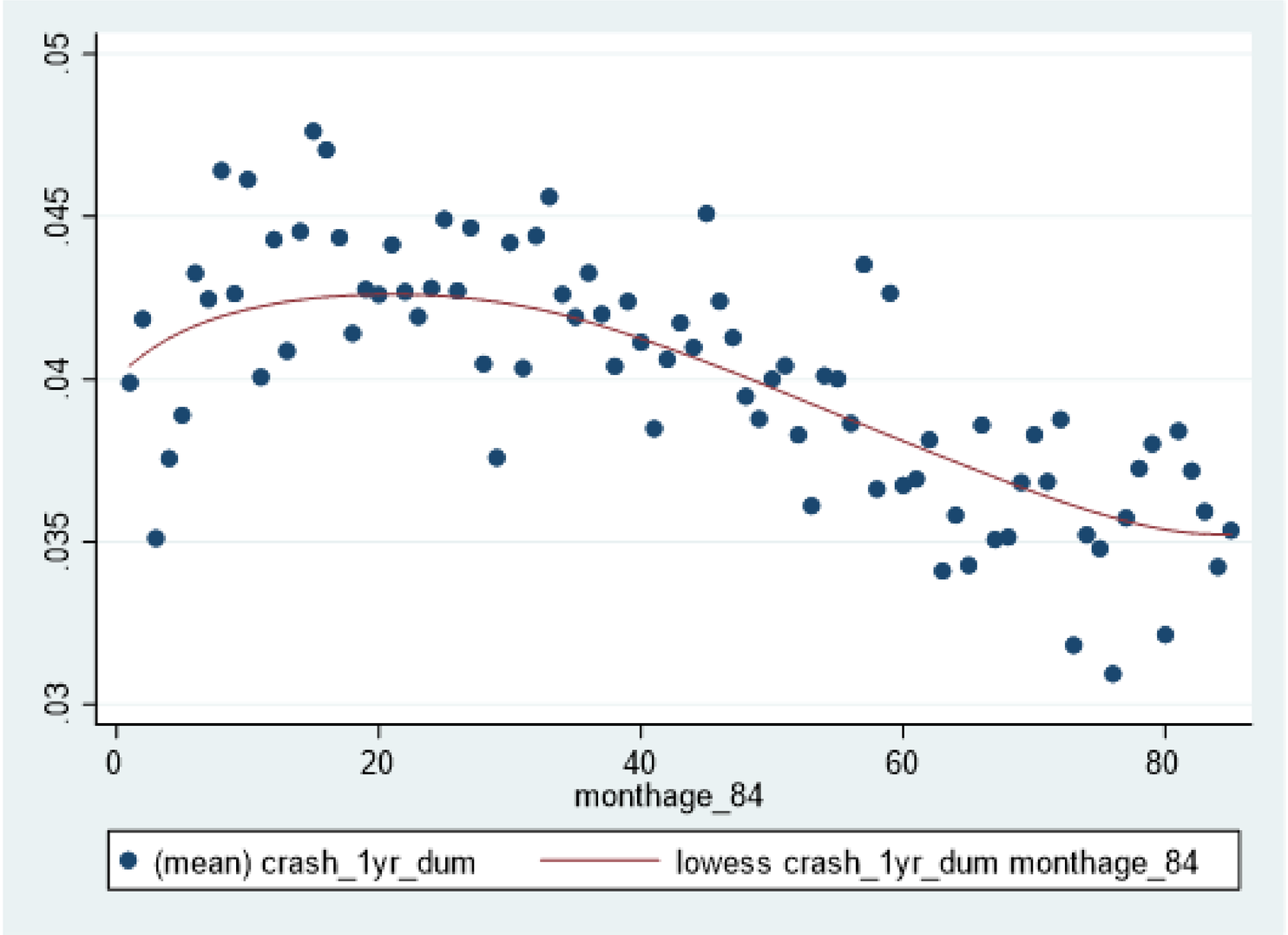
Our procedure

- There are total of 14 x 331 = 4,634 candidate models – i.e. 14 models with BW ranging from 35 to 365 days.
- We estimate the placebo treatment effect (which we know to be zero and constant across entities) using each candidate model. We repeat this for all **1826** placebo treatment thresholds, and assess the performance of each candidate model.
- Key stat is the **Root Mean Squared Error** (RMSE) of the estimated treatment effect = square root of the sum of the 1826 estimates
- Other stats are Coverage Rate (% of estimated CIs which include the true effect)

Extensions

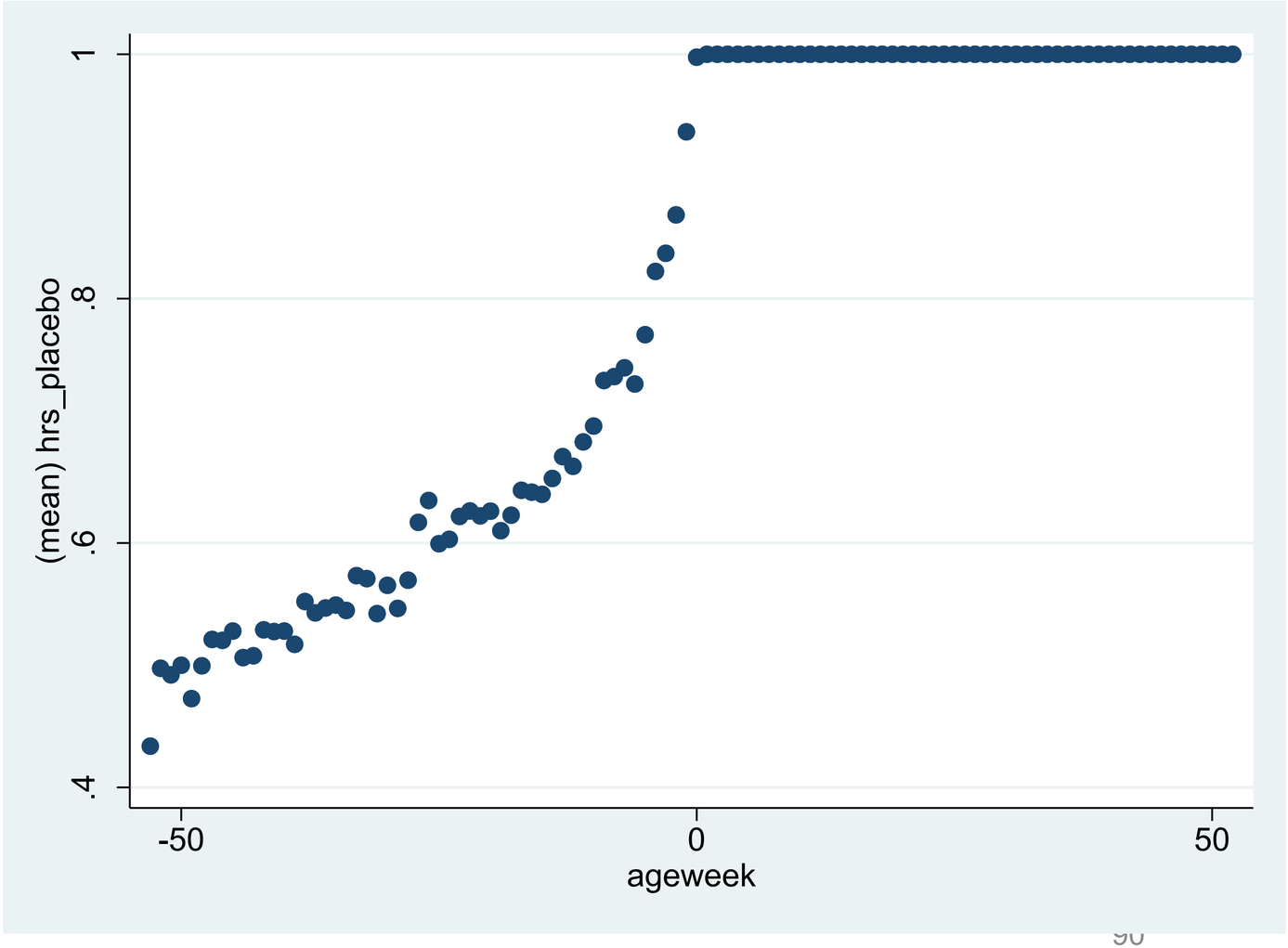
- Allow asymmetric bandwidths (to the left and right of the threshold) – these actually do better!
- *Impose* treatment effect heterogeneity into the placebo zone
  - Random perturbation of the real data for ‘treated’ observations
  - Marginal Treatment Effect which is linear in ‘resistance’

Outcome data in our placebo zone



Placebo treatments

Within this zone, we create placebo treatments in a way which mimics the true treatment selection process. For example, in the first placebo, persons are deemed treated if they obtained their license on or after 1 July 2001. The first stage relationship is shown.



Timing of Treatment Effect

	Best estimator	Best sym. cohort-IV	Best sym. RPJKD	Best sym. RKD	Best sym. RDD
	B: Timing of Treatment Effect				
MVA 6 months	-0.0094***	-0.0074**	-0.0078**	-0.0072	-0.0093**
SE	0.0031	0.0035	0.0036	0.0044	0.0044
MVA 6-12 months	-0.0048*	-0.0059*	-0.0070**	-0.0069*	-0.0077*
SE	0.0028	0.0034	0.0034	0.0040	0.0040
MVA 1-2 years	0.0035	0.0026	0.0027	0.0019	0.0033
SE	0.0036	0.0045	0.0046	0.0053	0.0053

Seriousness and Heterogeneity

	Best estimator	Best sym. cohort-IV	Best sym. RPJKD	Best sym. RKD	Best sym. RDD
	C: Serious MVAs				
Injury	-0.0084***	-0.0093***	-0.0100***	-0.0102***	-0.0110***
SE	0.0026	0.0032	0.0032	0.0038	0.0036
Fatality	-0.0002	-0.0001	-0.0002	-0.0002	-0.0002
SE	0.0003	0.0004	0.0004	0.0005	0.0005
	D: Heterogeneity by Sex				
MVA 1-year males	-0.0132**	-0.0146**	-0.0139*	-0.0114	-0.0163*
SE	0.0059	0.0072	0.0073	0.0086	0.0085
MVA 1-year females	-0.0164***	-0.0111*	-0.0159**	-0.0181**	-0.0177**
SE	0.0056	0.0066	0.0068	0.0080	0.0083

Other Applications

- We use our method to re-evaluate evidence on Head Start (Ludwig & Miller, 2007) and minimum legal drinking age (Lindo et al, 2016).
- In both cases, our method selects much larger BWs than the original studies.
- However, the conclusions of those studies are unchanged.