Estimation and Inference in RD Designs

Northwestern Causal Inference Workshop

Gonzalo Vazquez-Bare

Department of Economics, UC Santa Barbara

August 1, 2024

Estimation: overview

- Estimate regression functions at the left and right of the cutoff
- Global approach:
  - Estimate a p-th-order polynomial on full sample
  - Sensitive to misspecification
  - Erratic behavior at boundary points (Runge's phenomenon)
- "Flexible parametric" approach:
  - Estimate a polynomial within an ad-hoc bandwidth
- Local polynomial approach:
  - Data-driven bandwidth selection, nonparametric
  - Accounts for misspecification when performing inference

Global parametric approach

- Parametric assumption on conditional expectations, e.g.

$$\mathbb{E}[Y_i(d)|X_i] = \alpha_d + \beta_d(X_i - c)$$

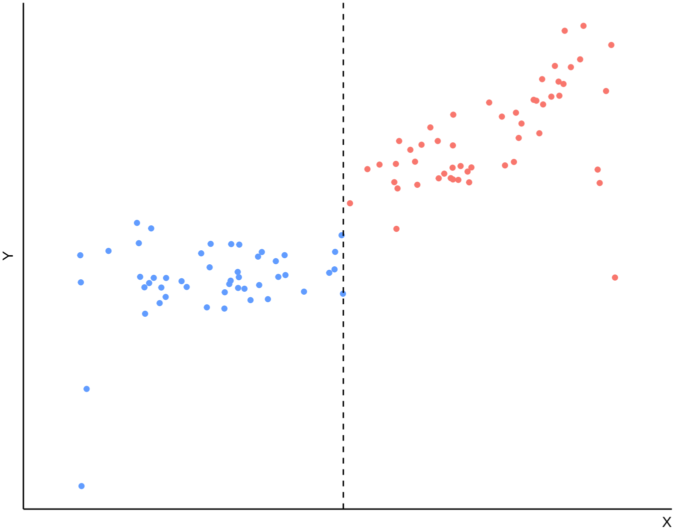
- This implies:

$$\mathbb{E}[Y_i|X_i] = \alpha + \beta D_i + \gamma(X_i - c) + \delta(X_i - c)D_i + u_i$$

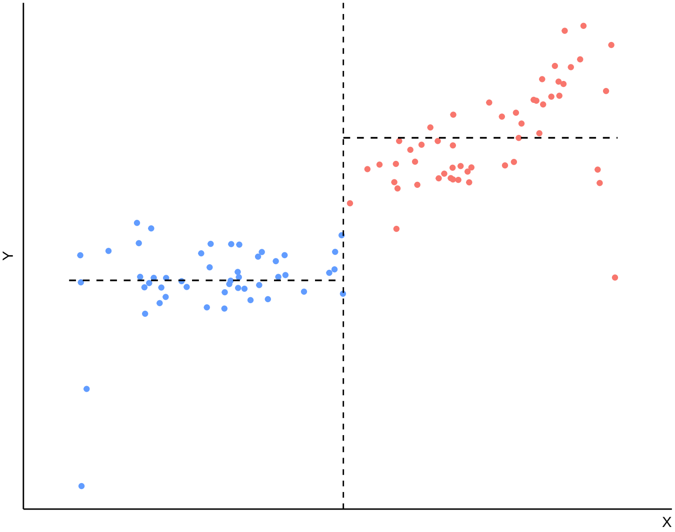
- Easily estimated by OLS on full sample

- The coefficient β recovers the treatment effect at the cutoff

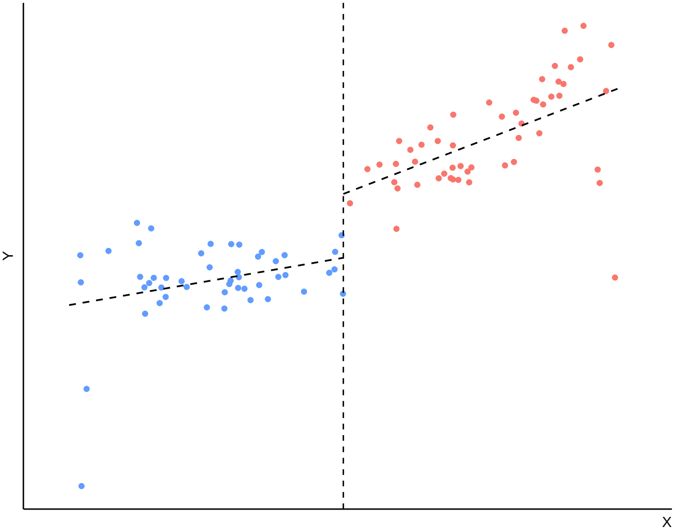
Global parametric approach



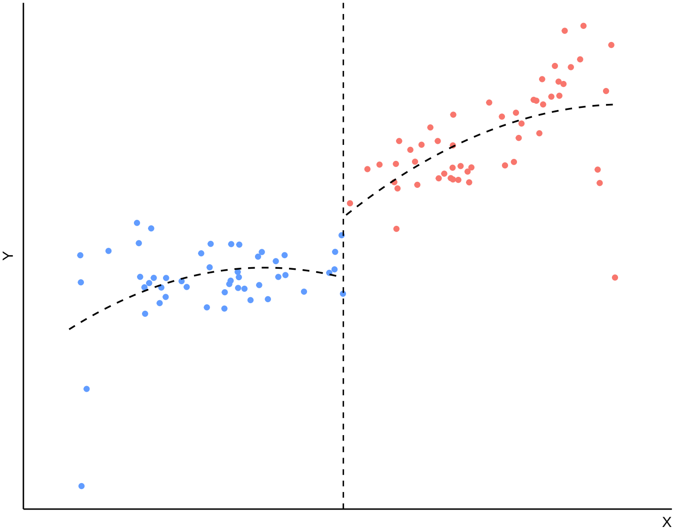
Global parametric approach: $p = 0$



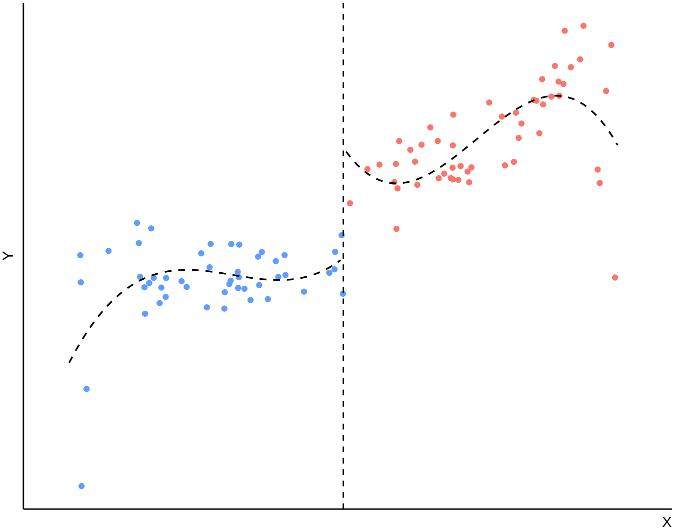
Global parametric approach: $p = 1$



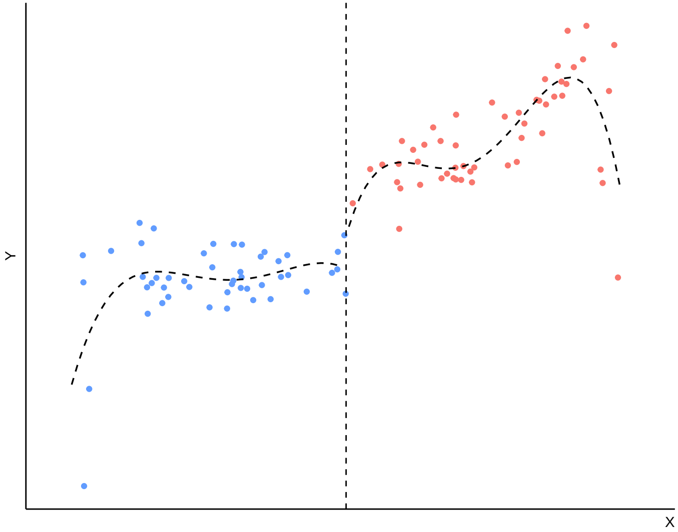
Global parametric approach: $p = 2$



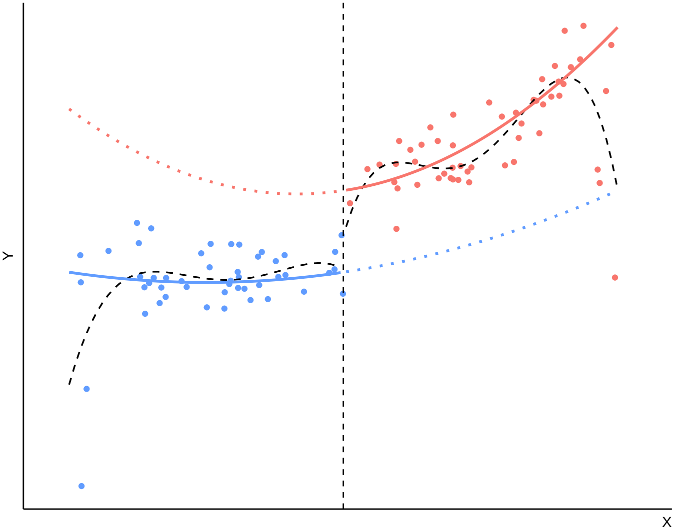
Global parametric approach: $p = 3$



Global parametric approach: $p = 4$



Global parametric approach: $p = 4$



Nonparametric estimation

- A misspecified parametric model yields biased estimators
  - Even with an infinitely large sample
- Nonparametric methods avoid bias due to model misspecification
  - “Let the data speak”
- RD effects can be estimated using local polynomial regression
  - Most common estimator: local linear regression

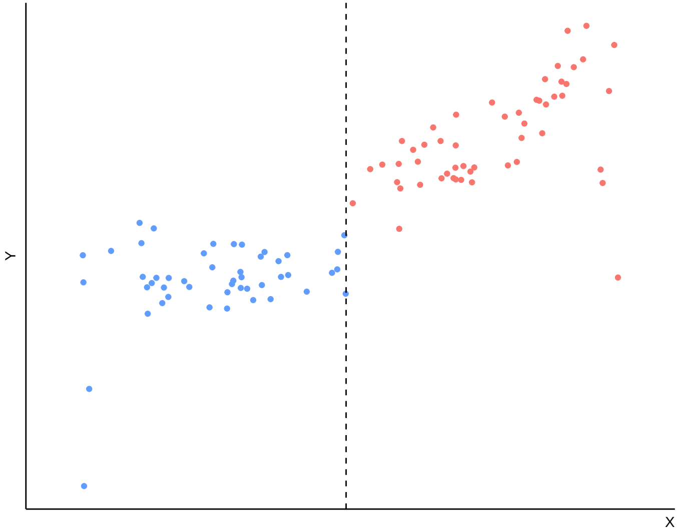
Local polynomial regression estimation

1. Choose the order of the polynomial p
2. Choose a weighting function (or *kernel*) $K(\cdot)$
3. Choose a bandwidth h
4. Estimate a regression of order p with weights $K(\cdot)$ in $[c-h,c)$
5. Estimate a regression of order p with weights $K(\cdot)$ in $[c,c+h]$
6. Construct the RD estimate:

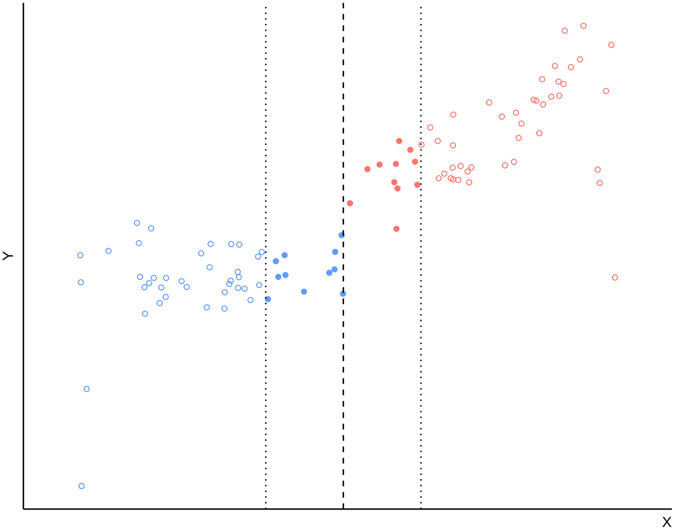
$$\hat{\tau} = \hat{\alpha}_+ - \hat{\alpha}_-$$

   where $\hat{\alpha}_+$ and $\hat{\alpha}_-$ are the regression intercepts

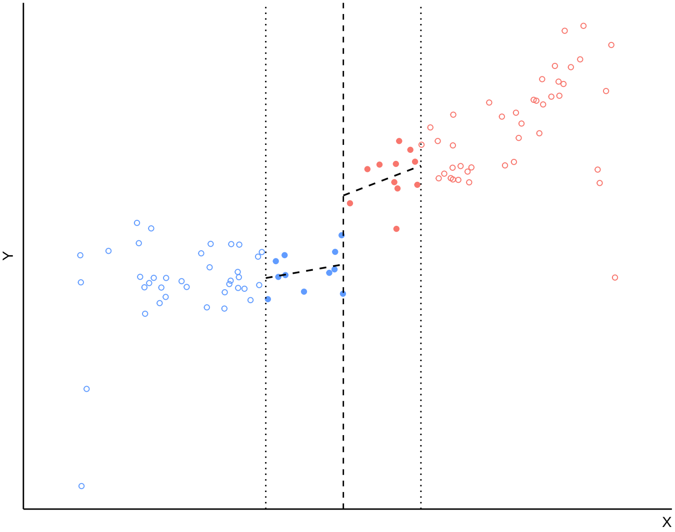
Local linear regression: estimation



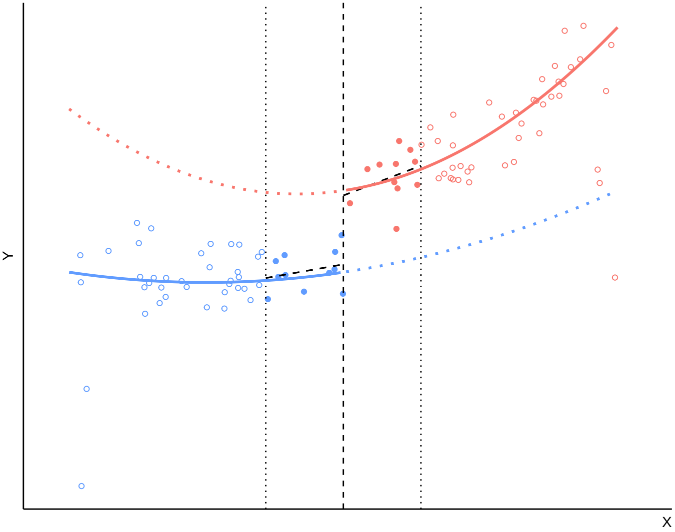
Local linear regression: estimation



Local linear regression: estimation



Local linear regression: estimation



Bandwidth selection

- Bias-variance trade-off in bandwidth selection
- A large bandwidth results in a large bias
  - The estimator uses observations far away from the cutoff
- A small bandwidth results in a large variance
  - The estimator uses few observations
- We can summarize this trade-off with the *mean squared error*:

$$MSE(\hat{\tau}) = Bias^2(\hat{\tau}) + Variance(\hat{\tau})$$

Bandwidth selection

- The *MSE-optimal bandwidth* is:

$$h^* = \arg\min_h MSE(\hat{\tau})$$

- It can be shown that:

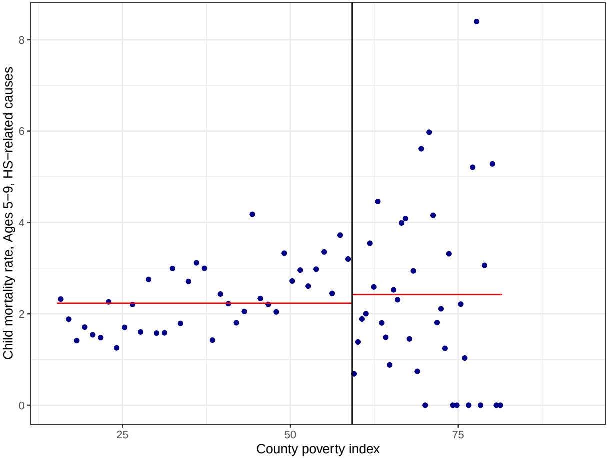
$$h^* = \left(\frac{V}{4B^2}\right)^{1/5} n^{-1/5}$$

- The MSE-optimal bandwidth optimally balances bias and variance
  - It adapts to the sample and the data generating process
  - If the bias is small relative to the variance, h^* will be large
  - The bandwidth gets smaller as the sample size n increases

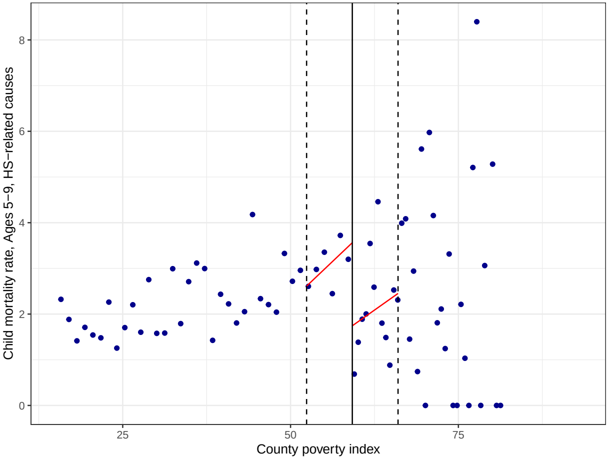
Other choices in local polynomial regression

- Order of the polynomial p
  - A larger p reduces the bias
  - But high-order polynomials have erratic behavior in finite samples
  - Common practice: use $p = 1$, try $p = 0$ or $p = 2$ for sensitivity
- Kernel (weighting function) $K(\cdot)$
  - Determines the weight of each observation within the bandwidth
  - Several options: uniform, triangular, Epanechnikov, normal,...
  - The triangular kernel is optimal for estimating boundary points
  - Typically this choice does not affect the results substantially

The effect of Head Start on child mortality



The effect of Head Start on child mortality



Inference in RD designs

- After estimating $\hat{\tau}$, we want to conduct inference
- Suppose we want to test $\mathsf{H}_0 : \tau = 0$ vs $\mathsf{H_A} : \tau \neq 0$
- The usual t-statistic:

$$t = \frac{\hat{\tau}}{se(\hat{\tau})}$$

  is biased

  - Since $h > 0$, we are using observations away from the cutoff

- We need to account for this bias in hypothesis testing

Inference in RD designs

- Use the bias-corrected t-statistic:

$$t_{\mathsf{bc}} = \frac{\hat{\tau} - \hat{B}}{se_{\mathsf{bc}}(\hat{\tau})}$$

  where $se_{\mathsf{bc}}(\hat{\tau})$ accounts for the estimation of the bias

- This is the *robust bias-corrected* approach to RD inference
  - Calonico, Cattaneo and Titiunik (2014)

- Implemented with the `rdrobust` package

  `https://rdpackages.github.io/rdrobust/`

Reporting empirical results

- Reporting the conventional point estimate and s.e. is not helpful
  - Conventional t-stat is biased, cannot be used to assess significance
- Reporting BC point estimate with RBC s.e. is suboptimal
  - Optimal bandwidth is calculated for the conventional estimate
- Recommended practice: report conventional point estimate with RBC p-value or CI

The effect of Head Start on child mortality

```
Sharp RD estimates using local polynomial regression.

Number of Obs.                 2783
BW type                       mserd
Kernel                   Triangular
VCE method                       NN

Number of Obs.                 2489          294
Eff. Number of Obs.             234          180
Order est. (p)                    1            1
Order bias  (q)                   2            2
BW est. (h)                   6.810        6.810
BW bias (b)                  10.725       10.725
rho (h/b)                     0.635        0.635
Unique Obs.                    2489          294

=============================================================================
        Method     Coef. Std. Err.         z     P>|z|      [ 95% C.I. ]
=============================================================================
  Conventional    -2.409     1.206    -1.998     0.046    [-4.772 , -0.046]
        Robust         -         -    -2.032     0.042    [-5.462 , -0.099]
=============================================================================
```

Assessing the Validity of RD Designs

Assessing the validity of RDDs

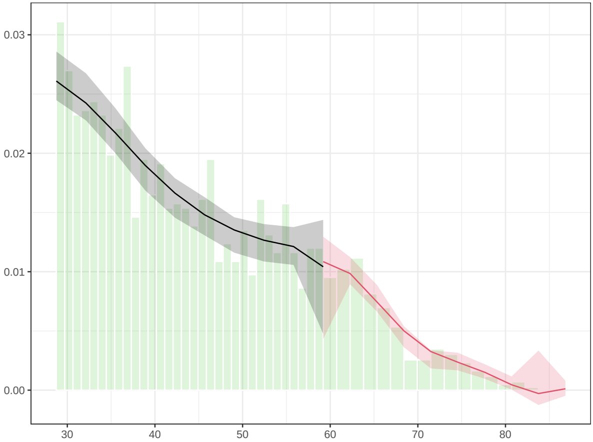
- The credibility of the results relies on the validity of the RD design
  - If the RD assumptions do not hold, the results are not valid
- The identification assumptions are untestable
  - They involve unobserved counterfactual magnitudes
- We discuss three ways to assess their credibility:
  1. Density discontinuity tests
  2. Continuity away from the cutoff
  3. Testing for discontinuities in covariates / placebo outcomes

Density discontinuity tests

- RDDs can be invalid if individuals manipulate X_i
  - E.g. if counties manipulate their poverty index to receive assistance
- Manipulation can imply sorting on one side of the cutoff
- Test wether the density of X_i is continuous around c
- McCrary (2008), Cattaneo, Jansson and Ma (2018)

  `https://rdpackages.github.io/rddensity/`

Head Start: density test



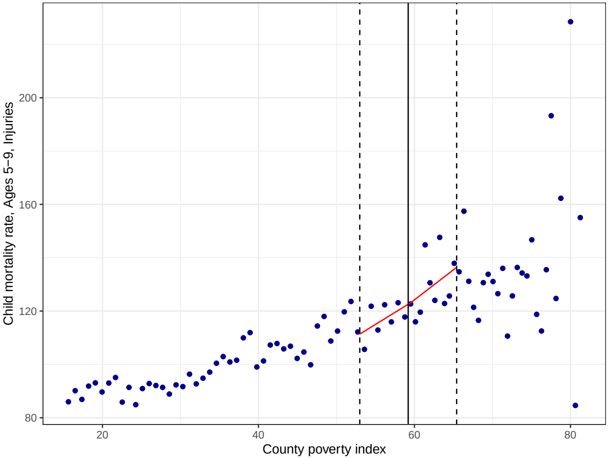
Continuity away from the cutoff

- Identification relies on continuity of $\mathbb{E}[Y_i(d)|X_i]$
- We can estimate $\mathbb{E}[Y_i(0)|X_i]$ below c, $\mathbb{E}[Y_i(1)|X_i]$ above c
- Check continuity away from the cutoff (graphically or formally)

Continuity in covariates / placebo outcomes

- Some variables should reveal no treatment effect:
  - Outcomes not targeted by treatment (*placebo outcomes*)
  - Exogenous or predetermined covariates
- Estimate an RD effect on these variables
- Finding a non-zero effect suggests an invalid RDD:
  - Existence of other (unobserved) treatments at the cutoff
  - Selection into treatment

Head Start: RD effect on placebo outcome (injuries)



Further issues and extensions

- Different bandwidths to the left and right of the cutoff
  - Arai and Ichimura (2015)
- Incorporating covariates
  - Calonico, Cattaneo, Farrell and Titiunik (2019)
- Coverage-error-rate (CER) optimal bandwidth
  - Calonico, Cattaneo and Farrell (2020)
- Inference for clustered data
  - Bartalotti and Brumett (2017)