Regression Discontinuity Design

Abi Adams

Hilary 2020

Overview

- Recap of RDD
- Kernel and Local Linear Regression in RDD
- Choice of Bandwidth
- Practicalities

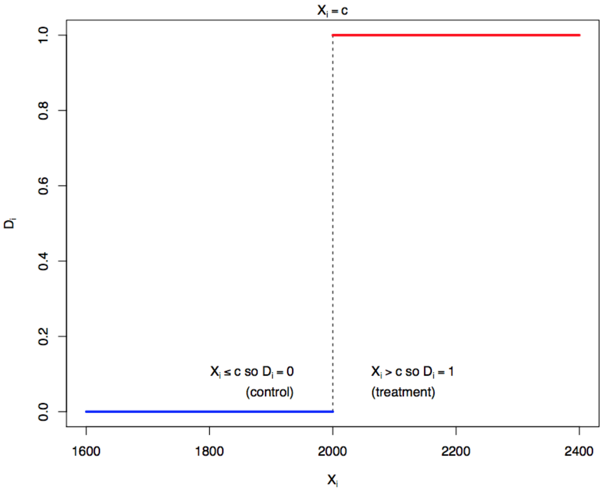
Sharp RDD: Basic Setup

- $D_i \in \{0, 1\}$: Treatment
- X_i: Forcing variable that determines the value of D_i at a cutpoint c

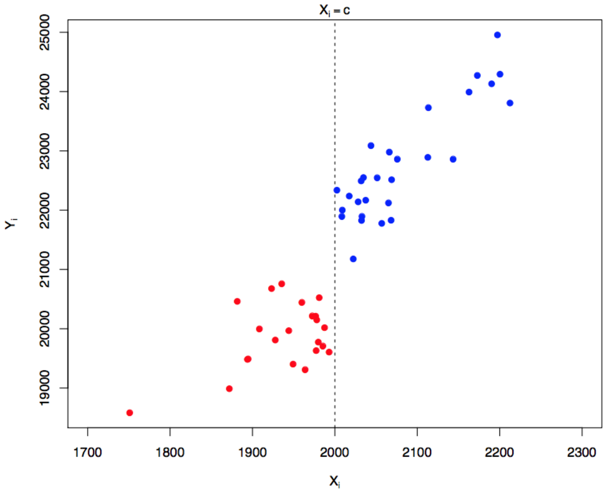
$$D_i = 1\left[X_i \geq c\right] \tag{1}$$

- Potential Outcomes framework
  - $Y_i(0)$: untreated potential outcome
  - $Y_i(1)$: treated potential outcome
  - $Y_i = (1 - D_i)Y_i(0) + D_i Y_i(1)$

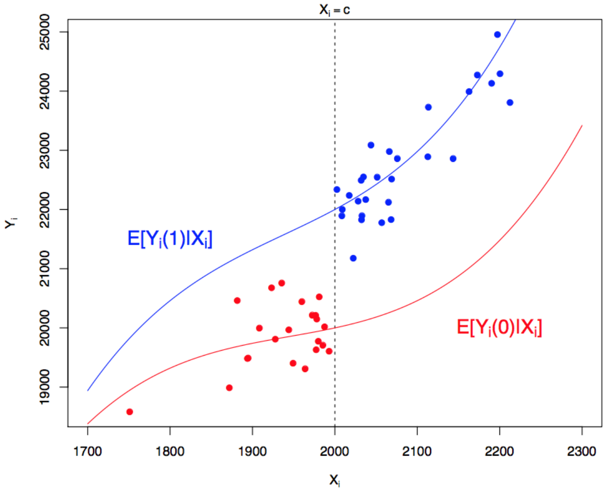
Probability of Treatment in Sharp RDD



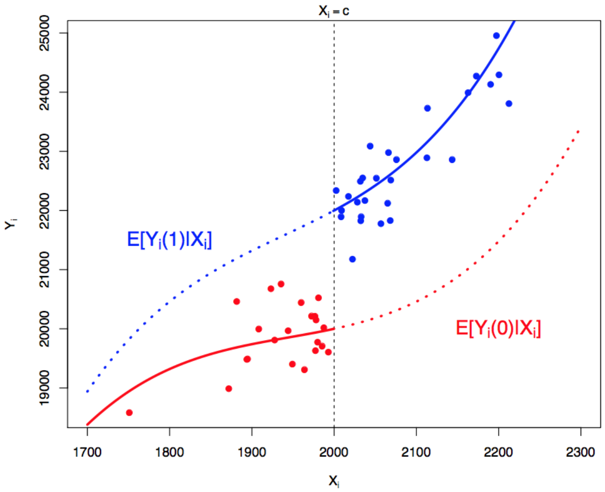
Graphical Illustration



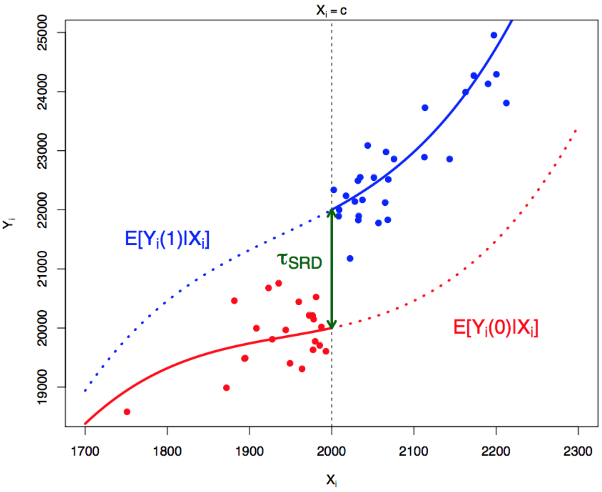
Graphical Illustration: Continuous $E[Y_i(d)|X_i]$



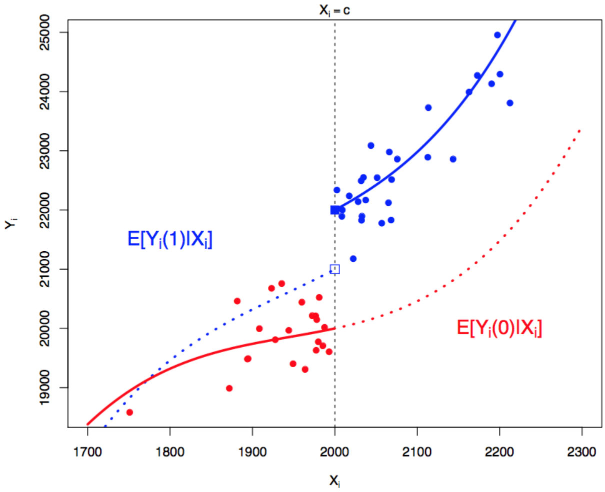
Graphical Illustration: Continuous $E[Y_i(d)|X_i]$



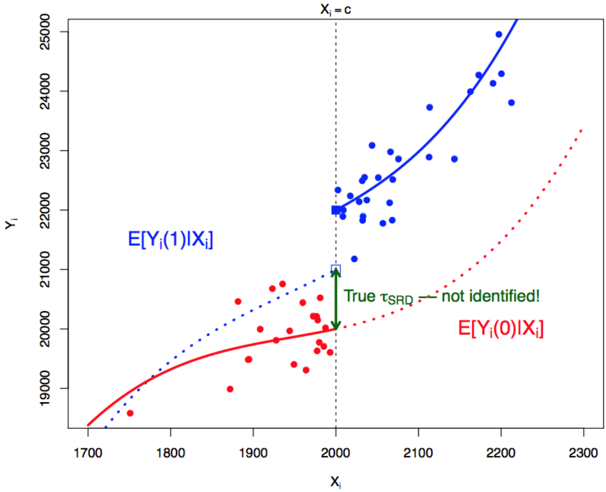
Graphical Illustration: Continuous $E[Y_i(d)|X_i]$



Graphical Illustration: Discontinuous $E[Y_i(d)|X_i]$



Graphical Illustration: Discontinuous $E[Y_i(d)|X_i]$



Identification of the Threshold Causal Effect

- Key assumption: continuity of average potential outcomes
- Casual estimand: Local ATE at the threshold

$$\tau_{RDD} = E\left[Y_i(1) - Y_i(0)|X_i = c\right] \tag{2}$$

- If the continuity assumption holds, then τ_{RDD} is identified as:

$$\tau_{RDD} = \lim_{x \downarrow c} E\left[Y_i|X_i = x\right] - \lim_{x \uparrow c} E\left[Y_i|X_i = x\right] \tag{3}$$

- Intuitively, continuity assumption allows one to do a little extrapolation to compensate for the lack of common support at the threshold

External Validity

- At best providing estimate of the treatment effect for the subpopulation with covariate equal to $X_i = c$
- With the so-called ‘fuzzy’ RDD, relevant subpopulation restricted even further to the ‘compliers’ at the threshold
- Whether this is an interesting causal effect of interest depends on the setting

Linear example

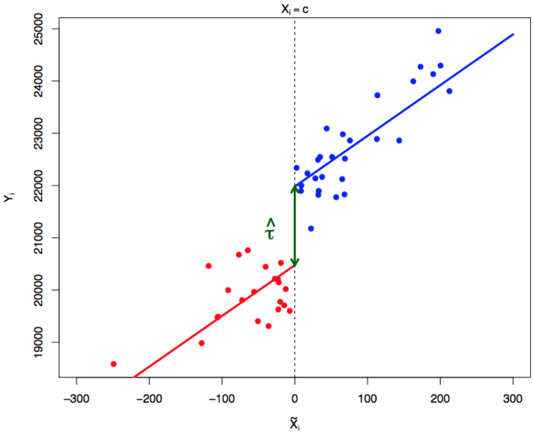
- Potential Outcomes

$$E[Y_i(0)|X_i] = \alpha + \beta X_i \tag{4}$$
$$E[Y_i(1)|X_i] = \tau + \alpha + \beta X_i \tag{5}$$
$$E[Y_i(1) - Y_i(0)|X_i] = \tau \tag{6}$$

- Can be estimated using a standard linear regression

$$E[Y_i|X_i, D_i] = (1 - D_i)E[Y_i(0)|X_i] + D_i E[Y_i(1)|X_i] \tag{7}$$
$$= \alpha + \tau D_i + \beta X_i \tag{8}$$
$$= \widetilde{\alpha} + \tau D_i + \beta \widetilde{X}_i \tag{9}$$

Linear Example



Linear example: Different Slopes

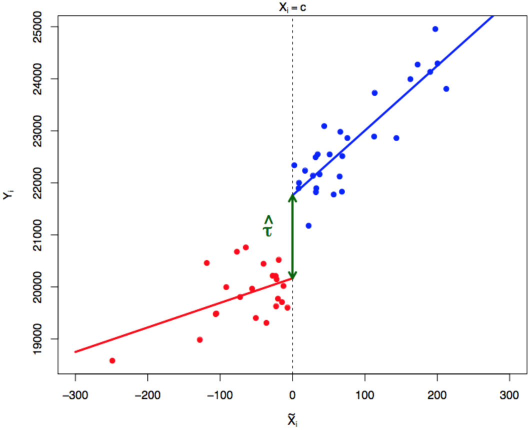
- Potential Outcomes

$$E[Y_i(0)|X_i] = \alpha_0 + \beta_0 X_i \quad (10)$$
$$E[Y_i(1)|X_i] = \alpha_1 + \beta_1 X_i \quad (11)$$
$$E[Y_i(1) - Y_i(0)|X_i] = \tau = (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)X_i \quad (12)$$

- Can be estimated using a standard linear regression

$$E[Y_i|X_i, D_i] = \widetilde{\alpha} + \beta_0 \widetilde{X}_i + \tau D_i + \widetilde{\beta} D_i \widetilde{X}_i \quad (13)$$

Linear Example: Different Slopes



Kernel Regression

- Could instead set up as a standard kernel regression to allow for non-linearities

$$\widehat{Y}_0(c) = \sum_{i:X_i<c} w_i(c)Y_i \tag{14}$$

$$\widehat{Y}_1(c) = \sum_{i:X_i\geq c} w_i(c)Y_i \tag{15}$$

- Where

$$w_i\left(c\right) = \frac{\frac{1}{nh}K\left(\frac{X_i-c}{h}\right)}{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i-c}{h}\right)}$$

Kernel Regression

- Consider a rectangular kernel as a special case

$$\widehat{\tau} = \frac{\sum Y_i \cdot 1\left[c \leq X_i \leq c+h\right]}{\sum 1\left[c \leq X_i \leq c+h\right]} - \frac{\sum Y_i \cdot 1\left[c-h \leq X_i < c\right]}{\sum 1\left[c-h \leq X_i < c\right]}$$

Simple Kernel Regression Issues

- Bias is high at the boundary and linear in the bandwidth (Porter, 2003)
- Often recommended to use local linear methods instead
- Some also advocate for the use of sieve methods (‘summing’) but the drawback here is that estimate of treatment effect can depend on values of the covariate far away from the cut off

Local Linear Regression

- As only care about estimating the regression function at a given point, can think of running mini regressions either side of the cutoff

$$\min_{\alpha_0,\beta_0} \sum_{i:c-h\leq X_i<c} (Y_i - \alpha_0 - \beta_0 \widetilde{X}_i)^2 \quad (16)$$

$$\min_{\alpha_1,\beta_1} \sum_{i:c\leq X_i\leq c+h} (Y_i - \alpha_1 - \beta_1 \widetilde{X}_i)^2 \quad (17)$$

Local Linear Regression

- Allows us to form the estimates of potential outcomes and treatment effects as follows

$$\widehat{Y}_0(c) = \widehat{\alpha}_0 + \widehat{\beta}_0(c - c) = \widehat{\alpha}_0 \tag{18}$$

$$\widehat{Y}_1(c) = \widehat{\alpha}_1 + \widehat{\beta}_1(c - c) = \widehat{\alpha}_1 \tag{19}$$

$$\widehat{\tau} = \widehat{\alpha}_1 - \widehat{\alpha}_0 \tag{20}$$

Picking the Bandwidth

- Best practise to use cross-validation
- Define cross-validation criterion as:

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}^{CV}(X_i))^2 \tag{21}$$

where

$$\widehat{Y}^{CV}(X_i) = \widehat{Y}_0^{CV}(X_i) \qquad \text{if } X_i < c \tag{22}$$

$$\widehat{Y}^{CV}(X_i) = \widehat{Y}_1^{CV}(X_i) \qquad \text{if } X_i \geq c \tag{23}$$

$$\tag{24}$$

and we use a 'one sided' kernel for estimation of $\widehat{\mu}$.

Picking the Bandwidth

- The precise window used for cross-validation gets rid of the interpolation problem and mimics behaviour at the boundaries

$$\min_{\alpha_0^{CV}(x),\beta_0^{CV}(x)} \sum_{j|x-h<X_j<x} (Y_i - \alpha_0^{CV} - \beta_0^{CV}(X_j - x))^2 \quad (25)$$

$$\min_{\alpha_1^{CV}(x),\beta_1^{CV}(x)} \sum_{j|x<X_j<x+h} (Y_i - \alpha_1^{CV} - \beta_1^{CV}(X_j - x))^2 \quad (26)$$

$$(27)$$

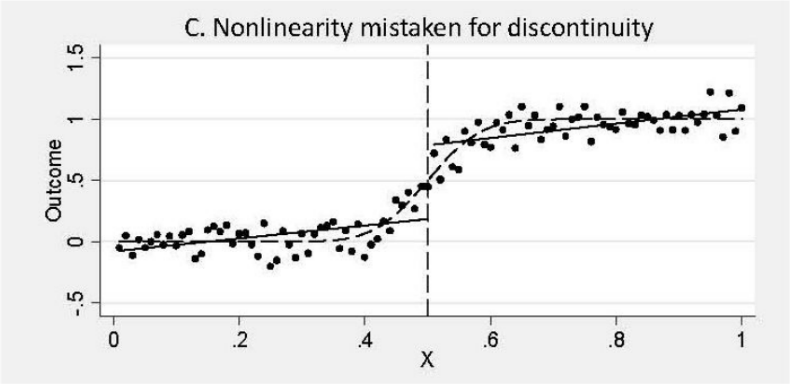
- We then pick the bandwidth to minimise the cross validation measure

$$h_{CV}^{opt} = arg\min_h CV(h) \quad (28)$$

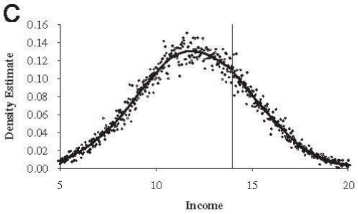
Diagnostics

- Explore robustness of results using falsification test
  - Sensitivity to alternative specifications
  - Balance tests: do any covariates jump at the threshold? (replace other covariates for Y)
  - Do jumps occur at placebo thresholds? (run with different values of c)
  - Is there any evidence of sorting around the threshold?

Specification Sensitivity



Sorting



C
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Density Estimate
5
10
15
20
Income
D
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Density Estimate
5
10
15
20
Income

Conclusion

- Aim has been to give a flavour of nonparametric and semiparametric regression
- Build familiarity so you can explore the methods in your own research
- What do I need to know for the exam?