Econometric Analysis of Cross Section and Panel Data

Lecture 10: Regression Discontinuity Design

Zhian Hu

Central University of Finance and Economics

Fall 2024

This Lecture

- Imbens, G. W. and T. Lemieus, 2008, "Regression Discontinuity Designs: A Guide to Practice", *Journal of Econometrics* 142: 615-635.

Introduction

- Since the late 1990s there has been a large number of studies in economics applying and extending regression discontinuity (RD) methods.
- In this lecture, we review some of the practical issues in implementation of RD methods.

Sharp and Fuzzy RD Designs: Basics

- We are interested in the causal effect of a **binary** intervention or treatment.
- Let $Y_i(0)$ and $Y_i(1)$ denote the pair of potential outcomes for unit i
  - $Y_i(0)$ is the outcome without exposure to the treatment
  - $Y_i(1)$ is the outcome given exposure to the treatment.
- Interest is the difference $Y_i(1) - Y_i(0)$, which is potentially heterogeneous across units.
- The fundamental problem of causal inference is that we never observe the pair $Y_i(1)$ and $Y_i(0)$ together.
- We therefore typically focus on average effects of the treatment, that is, averages of $Y_i(1) - Y_i(0)$ over (sub)populations, rather than on unit-level effects.

Sharp and Fuzzy RD Designs: Basics

- For each unit we observe the quadruple (Y_i, W_i, X_i, Z_i).
- Let $W_i \in \{0, 1\}$ denote the treatment received, with $W_i = 0$ if unit i was not exposed to the treatment, and $W_i = 1$ otherwise.
- The outcome observed can then be written as

$$Y_i = (1 - W_i) \cdot Y_i(0) + W_i \cdot Y_i(1) = \begin{cases} Y_i(0) & \text{if } W_i = 0 \\ Y_i(1) & \text{if } W_i = 1 \end{cases}$$

- In addition, we may observe a vector of covariates denoted by (X_i, Z_i), where X_i is a scalar and Z_i is an M-vector.
- X_i and Z_i are known not to have been affected by the treatment.

Sharp and Fuzzy RD Designs: Treatment Assignment

- Assignment to the treatment is determined, either completely or partly, by the value of X_i being on either side of a fixed threshold.
- This predictor may itself be associated with the potential outcomes, but this association is assumed to be smooth.
- So any discontinuity of the conditional distribution (or the conditional expectation) of the outcome as a function of this covariate at the cutoff value is interpreted as evidence of a causal effect of the treatment.

Sharp and Fuzzy RD Designs: A Few Settings

- Hahn et al. (1999) study the effect of an anti-discrimination law that only applies to firms with at least 15 employees.
- Matsudaira (2007) studies the effect of a remedial summer school program that is mandatory for students who score less than some cutoff level on a test.
- Eligibility for medical services through medicare is restricted by age (Card et al., 2004).
- Retirement, elections...

The Sharp Regression Discontinuity Design

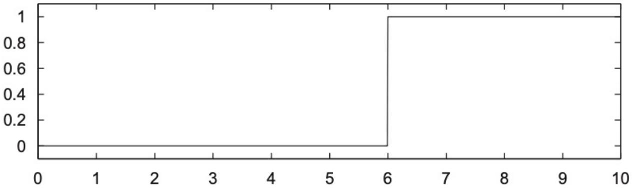


Fig. 1. Assignment probabilities (SRD).

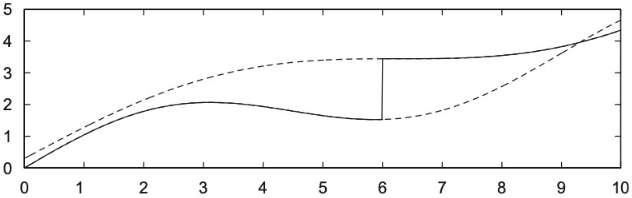


Fig. 2. Potential and observed outcome regression functions.

The Sharp Regression Discontinuity Design

- In the SRD design the assignment W_i is a deterministic function of one of the covariates, the forcing (or treatment-determining) variable X_i :

$$W_i = 1\{X_i \geqslant c\}.$$

- All units with a covariate value of at least c are assigned to the treatment group (and participation is mandatory for these individuals)
- All units with a covariate value less than c are assigned to the control group (members of this group are not eligible for the treatment).

The Sharp Regression Discontinuity Design

- Fig. 1 presents the conditional probability of receiving the treatment, $\Pr(W = 1 \mid X = x)$ against the covariate x. At $x = 6$ the probability jumps from 0 to 1 .
- In Fig. 2, three conditional expectations are plotted. The two continuous lines (partly dashed, partly solid) in the figure are the conditional expectations of the two **potential outcomes** given the covariate, $\mu_w(x) = \mathbb{E}[Y(w) \mid X = x]$, for $w = 0, 1$.
  - These two conditional expectations are continuous functions of the covariate.
  - We can only estimate $\mu_0(x)$ for $x < c$ and $\mu_1(x)$ for $x \geqslant c$.
- We plot the conditional expectation of the observed outcome, indicated by a solid line

$$\begin{aligned}\mathbb{E}[Y \mid X = x] = {} & \mathbb{E}[Y \mid W = 0, X = x] \cdot \Pr(W = 0 \mid X = x) \\ & + \mathbb{E}[Y \mid W = 1, X = x] \cdot \Pr(W = 1 \mid X = x)\end{aligned}$$

- The conditional expectation of the observed outcome jumps at $x = c = 6$.

The Sharp Regression Discontinuity Design

- In the SRD design we look at the discontinuity in the conditional expectation of the outcome given the covariate to uncover an average causal effect of the treatment:

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

which is interpreted as the average causal effect of the treatment at the discontinuity point

$$\tau_{\text{SRD}} = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = c\right].$$

The Sharp Regression Discontinuity Design: Assumptions

Classical identification assumptions include

1. Conditional Unconfoundedness Assumption:

$$Y_i(0), Y_i(1) \perp W_i \mid X_i$$

2. Overlap Assumption: for all values of the covariates there are both treated and control units, or

$$0 < \Pr\left(W_i = 1 \mid X_i = x\right) < 1$$

In the SRD design,

- The conditional unconfoundedness assumption holds in a trivial manner, because conditional on the covariates there is no variation in the treatment.
- **The Overlap Assumption is fundamentally violated.**

The Sharp Regression Discontinuity Design: Assumptions

- The lack of overlap, however, is compensated by the following continuity assumption.
- **Continuity of Conditional Regression Functions Assumption:** $\mathbb{E}[Y(0) \mid X = x]$ and $\mathbb{E}[Y(1) \mid X = x]$ are continuous in x.
- Under this assumption,

$$\mathbb{E}[Y(0) \mid X = c] = \lim_{x \uparrow c} \mathbb{E}[Y(0) \mid X = x] = \lim_{x \uparrow c} \mathbb{E}[Y \mid X = x]$$

$$\mathbb{E}[Y(1) \mid X = c] = \lim_{x \downarrow c} \mathbb{E}[Y(1) \mid X = x] = \lim_{x \downarrow c} \mathbb{E}[Y \mid X = x]$$

Thus, the average treatment effect at c, τ_{SRD}, satisfies

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

The Fuzzy Regression Discontinuity Design

- In the FRD design, the probability of receiving the treatment needs not change from 0 to 1 at the threshold.
- Instead, the design allows for a smaller jump in the probability of assignment to the treatment at the threshold:

$$\lim_{x \downarrow c} \Pr\left(W_i = 1 \mid X_i = x\right) \neq \lim_{x \uparrow c} \Pr\left(W_i = 1 \mid X_i = x\right)$$

  without requiring the jump to equal 1.
- Such a situation can arise if incentives to participate in a program change discontinuously at a threshold, without these incentives being powerful enough to move all units from nonparticipation to participation.

The Fuzzy Regression Discontinuity Design

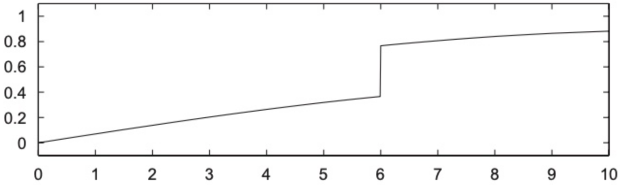


Fig. 3. Assignment probabilities (FRD).

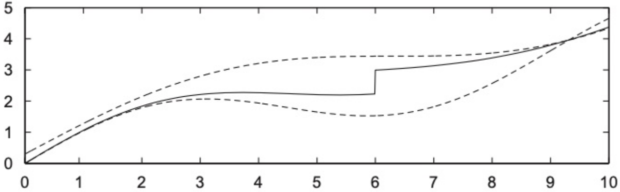


Fig. 4. Potential and observed outcome regression (FRD).

The Fuzzy Regression Discontinuity Design

- In this design we interpret the ratio of the jump in the regression of the outcome on the covariate to the jump in the regression of the treatment indicator on the covariate as an average causal effect of the treatment.

$$\tau_{\text{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y \mid X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \mid X = x]}{\lim_{x \downarrow c} \mathbb{E}[W \mid X = x] - \lim_{x \uparrow c} \mathbb{E}[W \mid X = x]}.$$

Fuzzy RD is IV

- Hahn, Todd and van der Klaauw (2001) exploit the instrumental variables connection to interpret the FRD design when the effect of the treatment varies by unit.
- A complier is a unit such that

$$\lim_{x \downarrow X_i} W_i(x) = 0 \quad \textit{and} \quad \lim_{x \uparrow X_i} W_i(x) = 1.$$

  - Compliers are units that would get the treatment if the cutoff were at X_i or below, but that would not get the treatment if the cutoff were higher than X_i.
- Never-takers are units with

$$\lim_{x \downarrow X_i} W_i(x) = 0 \quad \text{and} \quad \lim_{x \uparrow X_i} W_i(x) = 0,$$

  and always-takers are units with

$$\lim_{x \downarrow X_i} W_i(x) = 1 \quad \text{and} \quad \lim_{x \uparrow X_i} W_i(x) = 1.$$

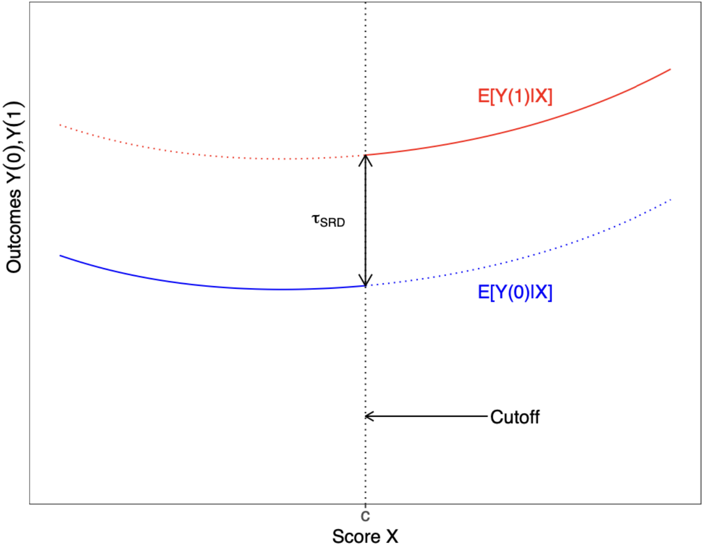
Fuzzy RD is IV

- Then,
$$\tau_{\text{FRD}} = \frac{\lim_{x \downarrow c} \mathbb{E}[Y \mid X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \mid X = x]}{\lim_{x \downarrow c} \mathbb{E}[W \mid X = x] - \lim_{x \uparrow c} \mathrm{E}\left[W \mid X = x\right]}$$
$$= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid \text{ unit } i \text{ is a complier and } X_i = c\right].$$
- The estimand is an average effect of the treatment, but only averaged for units with $X_i = c$ (by RD), and only for compliers (people who are affected by the threshold).

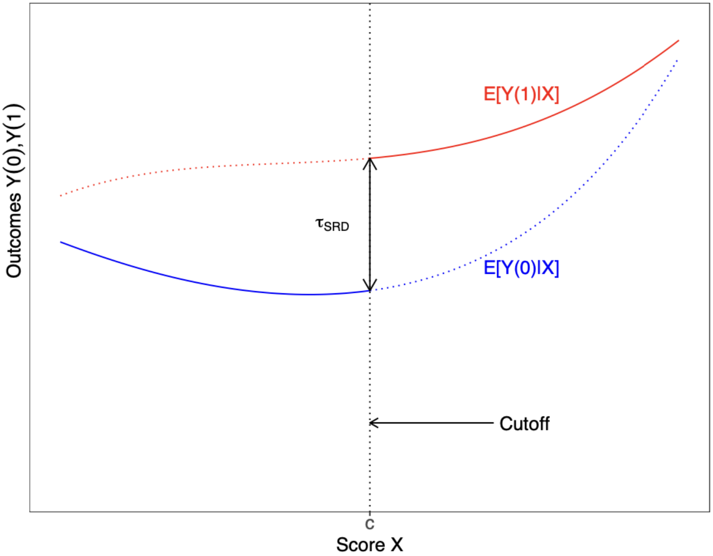
External Validity

- One important aspect of both the SRD and FRD designs is that they, at best, provide estimates of the average effect for a subpopulation, namely the subpopulation with covariate value equal to $X_i = c$.
- The FRD design restricts the relevant subpopulation even further to that of compliers at this value of the covariate.

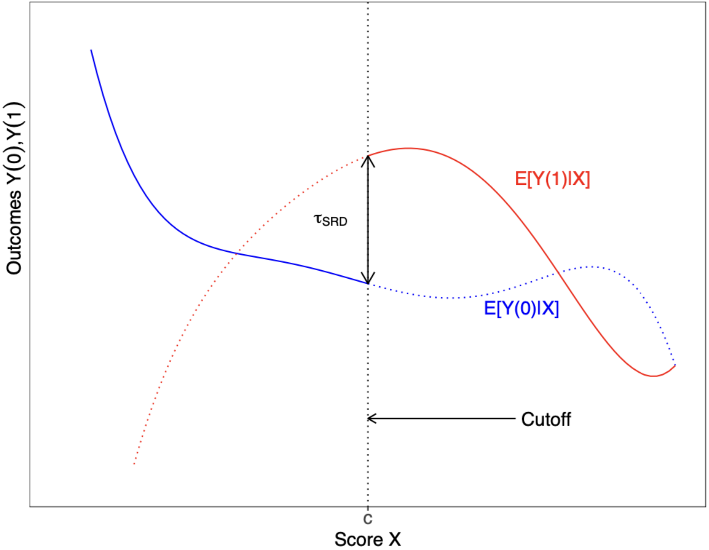
External Validity - No Heterogeneity



External Validity - Mild Heterogeneity



External Validity - Strong Heterogeneity



Graphical Analyses

- Graphical analyses should be an integral part of any RD analysis.
- The nature of RD designs suggests that the effect of the treatment of interest can be measured by the value of the discontinuity in the expected value of the outcome at a particular point.
- Inspecting the estimated version of this conditional expectation is a simple yet powerful way to visualize the identification strategy.

Graphical Analyses: Outcomes by Forcing Variable

- The first plot is a histogram-type estimate of the average value of the outcome for different values of the forcing variable, the estimated counterpart to the solid line in above figures.
- For some binwidth h, and for some number of bins K_0 and K_1 to the left and right of the cutoff value, respectively, construct bins $(b_k, b_{k+1}]$, for $k = 1, \ldots, K = K_0 + K_1$, where

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

  Then calculate the number of observations in each bin

$$N_k = \sum_{i=1}^{N} 1\left\{b_k < X_i \leqslant b_{k+1}\right\}$$

  and the average outcome in the bin

$$\bar{Y}_k = \frac{1}{N_k} \cdot \sum_{i=1}^{N} Y_i \cdot 1\left\{b_k < X_i \leqslant b_{k+1}\right\}.$$

Graphical Analyses: Outcomes by Forcing Variable

- The first plot of interest is that of the $\bar{Y}_k$, for $k = 1, \ldots, K$ against the mid point of the bins, $\tilde{b}_k = (b_k + b_{k+1})/2$.
- The question is whether around the threshold c there is any evidence of a jump in the conditional mean of the outcome.
- If the basic plot does not show any evidence of a discontinuity, there is relatively little chance that the more sophisticated analyses will lead to robust and credible estimates with statistically and substantially significant magnitudes.
- Any other jumps at points which you are not interested in?

Graphical Analyses: Covariates by Forcing Variable

- The second set of plots compares average values of other covariates in the K bins. Specifically, let Z_i be the M-vector of additional covariates, with mth element Z_{im}. Then calculate

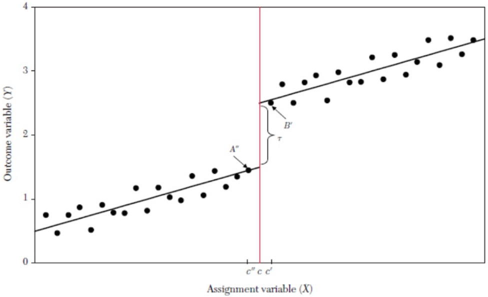
$$\bar{Z}_{km} = \frac{1}{N_k} \cdot \sum_{i=1}^{N} Z_{im} \cdot 1\left\{b_k < X_i \leqslant b_{k+1}\right\}.$$

  The second plot of interest is that of the $\bar{Z}_{km}$, for $k = 1, \ldots, K$ against the mid point of the bins, $\tilde{b}_k$, for all $m = 1, \ldots, M$.
- In the case of FRD designs, it is also particularly useful to plot the mean values of the treatment variable W_i to make sure there is indeed a jump in the probability of treatment at the cutoff point.

Graphical Analyses: The Density of the Forcing Variable

- In the third graph, one should plot the number of observations in each bin, N_k, against the mid points $\tilde{b}_k$.
- This plot can be used to inspect whether there is a discontinuity in the distribution of the forcing variable X at the threshold.
- Such discontinuity would raise the question of whether the value of this covariate was manipulated by the individual agent, invalidating the design.

Estimation



- Parametric estimation with a linear form (?):

$$Y = \alpha + D\tau + X\beta + \varepsilon$$

- A simple way of relaxing the linearity assumption is to include polynomial functions of X in the regression model.
- The main concern is that they are more sensitive to outcome values for observations far away from the cutoff point.

Estimation: Nonparametric Regression at the Boundary

- Kernel regression: First define the conditional means

$$\mu_l(x) = \lim_{z \uparrow x} \mathbb{E}[Y(0) \mid X = z] \quad \text{and} \quad \mu_r(x) = \lim_{z \downarrow x} \mathbb{E}[Y(1) \mid X = z]$$

- The estimand in the SRD design is, in terms of these regression functions,

$$\tau_{\mathrm{SRD}} = \mu_r(c) - \mu_l(c).$$

- A natural approach is to use standard nonparametric regression methods for estimation of $\mu_l(x)$ and $\mu_r(x)$.

Estimation: Nonparametric Regression at the Boundary

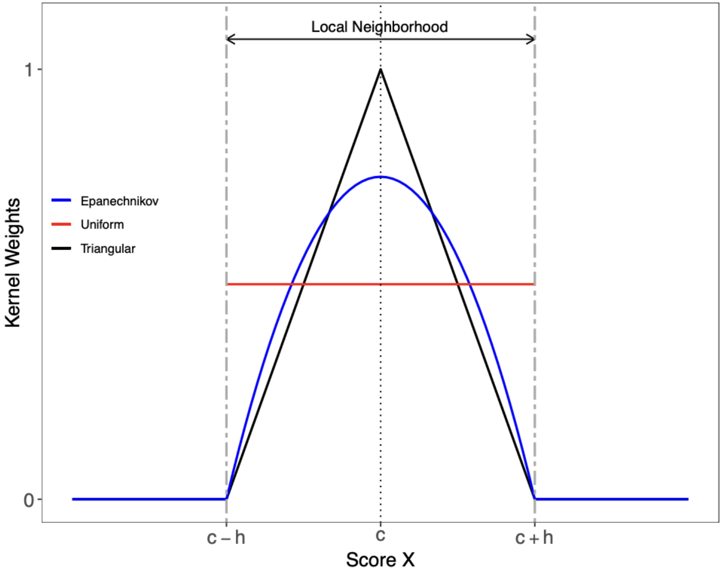
- Suppose we use a kernel $K(u)$, with $\int K(u)\mathrm{d}u = 1$. Then the regression functions at x can be estimated as

$$\hat{\mu}_{\mathrm{l}}(x) = \frac{\sum_{i:X_i<x} Y_i \cdot K\left((X_i - x)/h\right)}{\sum_{i:X_i<x} K\left((X_i - x)/h\right)}$$
$$\hat{\mu}_{\mathrm{r}}(x) = \frac{\sum_{i:X_i\geqslant x} Y_i \cdot K\left((X_i - x)/h\right)}{\sum_{i:X_i\geqslant x} K\left((X_i - x)/h\right)}$$

  where h is the bandwidth.
- The estimator for the object of interest is then

$$\begin{aligned}\hat{\tau}_{\mathrm{SRD}} &= \hat{\mu}_{\mathrm{r}}(c) - \hat{\mu}_{\mathrm{l}}(c) \\ &= \frac{\sum_{i:X_i\geqslant c} Y_i \cdot K\left((X_i - c)/h\right)}{\sum_{i:X_i\geqslant c} K\left((X_i - c)/h\right)} - \frac{\sum_{i:X_i<c} Y_i \cdot K\left((X_i - c)/h\right)}{\sum_{i:X_i<c} K\left((X_i - c)/h\right)}.\end{aligned}$$

Estimation: Nonparametric Regression at the Boundary



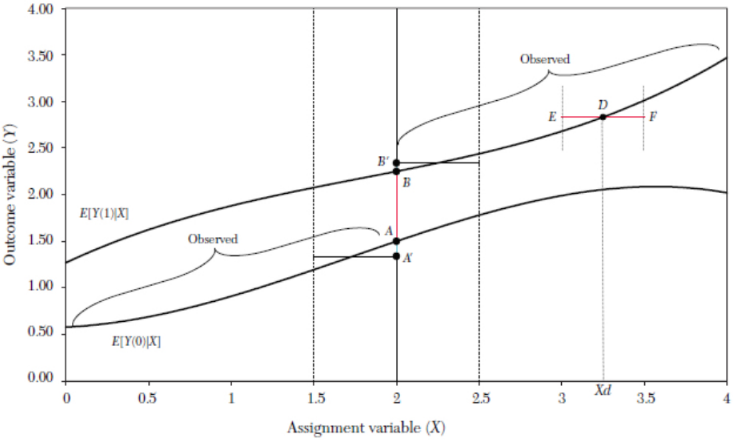
Estimation: Nonparametric Regression at the Boundary

- A special case: Suppose we use a uniform kernel, e.g., $K(u) = \frac{1}{2}$ for $-1 < u < 1$, and 0 elsewhere. Then the estimator can be written as

$$\begin{aligned}\hat{\tau}_{\mathrm{SRD}} &= \frac{\sum_{i=1}^{N_{hr}} Y_i \cdot 1\{c \leqslant X_i \leqslant c+h\}}{\sum_{i=1}^{N_{hr}} 1\{c \leqslant X_i \leqslant c+h\}} - \frac{\sum_{i=1}^{N_{lr}} Y_i \cdot 1\{c-h \leqslant X_i < c\}}{\sum_{i=1}^{N_{lr}} 1\{c-h \leqslant X_i < c\}} \\ &= \bar{Y}_{h\mathrm{r}} - \bar{Y}_{h1},\end{aligned}$$

  the difference between the average outcomes for observations within a distance h of the cutoff point on the right and left of the cutoff, respectively.
- This estimator can be interpreted as first discarding all observations with a value of X_i more than h away from the discontinuity point c, and then simply differencing the average outcomes by treatment status in the remaining sample.

Estimation: Nonparametric Regression at the Boundary



- The bias is linear in the bandwidth h.

Estimation: Local Linear Regression

- Instead of locally fitting a constant function, we can fit linear regression functions to the observations within a distance h on either side of the discontinuity point

$$\min_{\alpha_{\mathrm{l}}:\beta_{\mathrm{l}}} \sum_{i:c-h<X_i<c} \left(Y_i - \alpha_{\mathrm{l}} - \beta_{\mathrm{l}} \cdot (X_i - c)\right)^2$$

and

$$\min_{\alpha_{\mathrm{r}}:\beta_{\mathrm{r}}} \sum_{i:c\leqslant X_i<c+h} \left(Y_i - \alpha_{\mathrm{r}} - \beta_{\mathrm{r}} \cdot (X_i - c)\right)^2 .$$

Estimation: Local Linear Regression

- The value of $\mu_{\mathrm{l}}(c)$ is then estimated as

$$\widehat{\mu_{\mathrm{l}}(c)} = \hat{\alpha}_{\mathrm{l}} + \hat{\beta}_{\mathrm{l}} \cdot (c - c) = \hat{\alpha}_{\mathrm{l}},$$

  and the value of $\mu_{\mathrm{r}}(c)$ is then estimated as

$$\widehat{\mu_{\mathrm{r}}(c)} = \hat{\alpha}_{\mathrm{r}} + \hat{\beta}_{\mathrm{r}} \cdot (c - c) = \hat{\alpha}_{\mathrm{r}}.$$

- Given these estimates, the average treatment effect is estimated as

$$\hat{\tau}_{\mathrm{SRD}} = \hat{\alpha}_{\mathrm{r}} - \hat{\alpha}_{\mathrm{l}}.$$

- Alternatively one can estimate the average effect directly in a single regression, by solving

$$\min_{\alpha,\beta,\tau,\gamma} \sum_{i=1}^{N} 1\left\{c - h \leqslant X_i \leqslant c + h\right\} \cdot \left(Y_i - \alpha - \beta \cdot (X_i - c) - \tau \cdot W_i - \gamma \cdot (X_i - c) \cdot W_i\right)^2$$

  which will numerically yield the same estimate of τ_{SRD}.

Estimation: Local Linear Regression

- An alternative is to impose the restriction that the slope coefficients are the same on both sides of the discontinuity point. This can be imposed by requiring that $\beta_l = \beta_r$.
- We can make the nonparametric regression more sophisticated by using weights that decrease smoothly as the distance to the cutoff point increases, instead of the 0/1 weights based on the rectangular kernel.
- For inference we can use standard least squares methods.

Estimation: Local Linear Regression with Covariates

- If the conditional distribution of the covariates Z given X is continuous at $x = c$, the presence of these covariates rarely changes the identification strategy.
- Including additional covariates may eliminate some bias that is the result of the inclusion of observations with values of X not too close to c.
- The presence of the covariates can improve precision if Z is correlated with the potential outcomes.

Estimation for the FRD Design

- First, consider local linear regression for the outcome, on both sides of the discontinuity point. Let

$$\left(\hat{\alpha}_{yl}, \hat{\beta}_{yl}\right) = \arg\min_{\alpha_{yl},\beta_{yl}} \sum_{i: c-h \leqslant X_i < c} \left(Y_i - \alpha_{yl} - \beta_{yl} \cdot (X_i - c)\right)^2 ,$$

$$\left(\hat{\alpha}_{\mathrm{yr}}, \hat{\beta}_{\mathrm{yr}}\right) = \arg\min_{\alpha_{yr},\beta_{yr}} \sum_{i: c \leqslant X_i \leqslant c+h} \left(Y_i - \alpha_{yr} - \beta_{yr} \cdot (X_i - c)\right)^2 .$$

- The magnitude of the discontinuity in the outcome regression is then estimated as

$$\hat{\tau}_y = \hat{\alpha}_{\mathrm{yr}} - \hat{\alpha}_{yl}.$$

Estimation for the FRD Design

- Second, consider the two local linear regression for the treatment indicator:

$$\left(\hat{\alpha}_{wl}, \hat{\beta}_{wl}\right) = \arg\min_{\alpha_{wl}, \beta_{wl}} \sum_{i: c-h \leqslant X_i < c} \left(W_i - \alpha_{wl} - \beta_{wl} \cdot (X_i - c)\right)^2$$

$$\left(\hat{\alpha}_{wr}, \hat{\beta}_{wr}\right) = \arg\min_{\alpha_{wr}, \beta_{wr}} \sum_{i: c \leqslant X_i \leqslant c+h} \left(W_i - \alpha_{wr} - \beta_{wr} \cdot (X_i - c)\right)^2$$

- The magnitude of the discontinuity in the treatment regression is then estimated as

$$\hat{\tau}_w = \hat{\alpha}_{w\mathrm{r}} - \hat{\alpha}_{wl}.$$

- Finally, we estimate the effect of interest as the ratio of the two discontinuities:

$$\hat{\tau}_{\mathrm{FRD}} = \frac{\hat{\tau}_y}{\hat{\tau}_w} = \frac{\hat{\alpha}_{y\mathrm{r}} - \hat{\alpha}_{yl}}{\hat{\alpha}_{w\mathrm{r}} - \hat{\alpha}_{wl}}$$

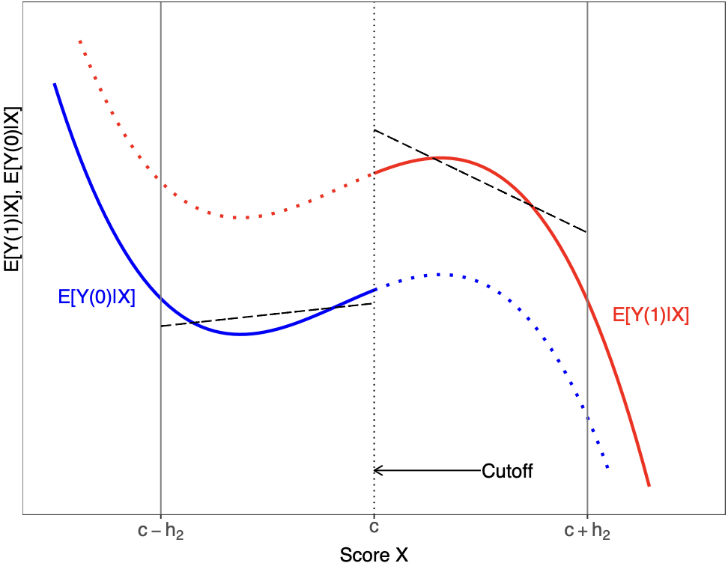
Estimation for the FRD Design: 2SLS

- Define
$$V_i = \begin{pmatrix} 1 \\ 1\{X_i < c\} \cdot (X_i - c) \\ 1\{X_i \geqslant c\} \cdot (X_i - c) \end{pmatrix} \text{ and } \delta = \begin{pmatrix} \alpha_{yl} \\ \beta_{yl} \\ \beta_{yr} \end{pmatrix}$$
- Then we can write
$$Y_i = \delta' V_i + \tau \cdot W_i + \varepsilon_i.$$
- Estimating τ based on the regression function by 2SLS methods, with the indicator $1\{X_i \geqslant c\}$ as the excluded instrument and V_i as the set of exogenous variables.
- It is numerically identical to $\hat{\tau}_{\mathrm{FRD}}$ as given in previous slides.

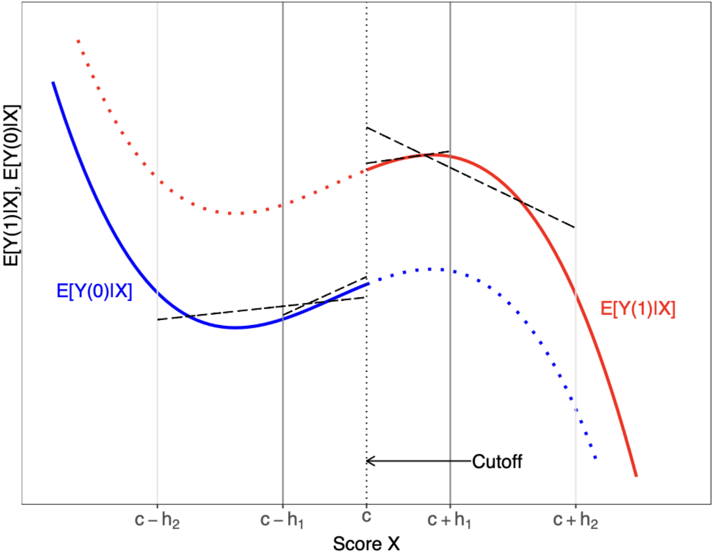
Bandwidth

- How to choose the bandwidth h?
- In general, choosing a bandwidth in nonparametric estimation involves finding an optimal balance between precision and bias.
- Using a larger bandwidth yields more precise estimates as more observations are available to estimate the regression.
- The linear specification is less likely to be accurate when a larger bandwidth is used, which can bias the estimate of the treatment effect.

Bandwidth



Bandwidth



Bandwidth

- Several methods to choose optimal bandwidth.
- Cross validation (Ludwig and Miller, 2007)

$$\mathrm{CV}_Y(h) = \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \hat{\mu}\left(X_i\right)\right)^2$$

$$h_{\mathrm{CV}}^{\mathrm{opt}} = \arg\min_h \mathrm{CV}_Y(h)$$

- Imbens and Kalyanaraman (2012, RES)
- Calonico, Cattaneo, and Titiunik (2014, ECMA)

Specification Testing

- Tests involving covariates
  - A first concern about RD designs is the possibility of other changes at the same cutoff value of the forcing variable.
  - Such changes may affect the outcome, and these effects may be attributed erroneously to the treatment of interest.
- Tests of continuity of the density
  - The second concern is that of manipulation of the forcing variable.
  - McCrary (2007) suggests testing the null hypothesis of continuity of the density of the forcing variable.
  - Testing for jumps at non-discontinuity points

A Guide to Practice: Sharp RD

- Graph the data by computing the average value of the outcome variable over a set of bins.
  - The binwidth has to be large enough to have a sufficient amount of precision so that the plots looks smooth on either side of the cutoff value.
  - but at the same time small enough to make the jump around the cutoff value clear.
- Estimate the treatment effect by running linear regressions on both sides of the cutoff point. Since we propose to use a rectangular kernel, these are just standard regression estimated within a bin of width h on both sides of the cutoff point. Note that:
  - Standard errors can be computed using standard least square methods (robust standard errors).
  - The optimal bandwidth can be chosen using cross-validation methods.

A Guide to Practice: Sharp RD

- The robustness of the results should be assessed by employing various specification tests.
  - Looking at possible jumps in the value of other covariates at the cutoff point
  - Testing for possible discontinuities in the conditional density of the forcing variable
  - Looking whether the average outcome is discontinuous at other values of the forcing variable
  - Using various values of the bandwidth, with and without other covariates that may be available.

A Guide to Practice: Fuzzy RD

- Graph the average outcomes over a set of bins as in the case of SRD, but also graph the probability of treatment.
- Estimate the treatment effect using TSLS, which is numerically equivalent to computing the ratio in the estimate of the jump (at the cutoff point) in the outcome variable over the jump in the treatment variable.
  - Standard errors can be computed using the usual (robust) 2SLS standard errors
  - The optimal bandwidth can again be chosen using a modified cross-validation procedure
- The robustness of the results can be assessed using the various specification tests mentioned in the case of SRD designs.