Regression Discontinuity Design
Econometric Issues

Brian P. McCall

University of Michigan

Texas Schools Project, University of Texas, Dallas
November 20, 2009

Regression Discontinuity Design
Outline of Talk

- Introduction
- What is the regression discontinuity design?
    - Sharp versus Fuzzy design
- What are the critical assumptions that we need to assess?

Introduction

- Regression Discontinuity Design (RDD) first implemented in 1960 by Thistlethwaite & Campbell in their study of National Merit Scholarship Program
- Recently seen a resurgence in economics to study such diverse topics as:
  - Class size on test scores (Angrist & Levy, 1999).
  - Extended benefit receipt on unemployment durations (Card et al., 2007).
  - Financial aid on College Attendance (Kane, 2003).
  - Union victory in NLRB election on wages (DiNardo & Lee, 2004).

Introduction

  - Attainment of minimum drinking age on mortality (Carpenter & Dobkin, 2009).
  - Effect of remediation on college outcomes (McFarlin, 2009)
- Why is it so popular?
  - Under fairly general conditions (most of which are testable) it allows researcher to make causal statements regarding impact of treatment on outcome.

Sharp RDD

- Notation & Definitions:
  - D– dummy variable equal to one if individual receives treatment and 0 if individual doesn't receive treatment.
  - Y – the outcome variable.
  - X – the running variable or assignment variable.

- Definition: A **sharp** regression discontinuity design is such that D = 1 if and only if X □c where c is referred to as the "cut-point".

Sharp RDD

- Lets look at the simple model

  Y = □+ □*D* + □*X* + □

- Suppose that *X* was randomly assigned to individuals and that *D* = 1 if and only if *X* □*c*.
- Then we have a randomized design and we can estimate □by:

$$\hat{\tau} = \frac{\sum_{i=1}^{n} D_i Y_i}{\sum_{i=1}^{n} D_i} - \frac{\sum_{i=1}^{n} (1 - D_i) Y_i}{\sum_{i=1}^{n} (1 - D_i)}$$

Sharp RDD

- What if X is not randomly assigned?
- **<u>Assumption:</u>** E(□|X) is a continuous function of X.
- Let Y_1 be the value of Y if treatment is received and Y_0 be the value of Y if no treatment is received. So
- $E(Y_1|X=c + d) = \square + \square + \square X + E(\square|X=c + d)$
- $E(Y_0|X=c - d) = \square + \square X + E(\square|X=c - d)$

Sharp RDD

- $E(Y_1|X=c+d) - E(Y_0|X=c+d) = \square +$ $E(\square|X=c+d) - E(\square|X=c-d)$
- Taking limits as d goes to 0 then by the continuity of E($\square$|X) we have:

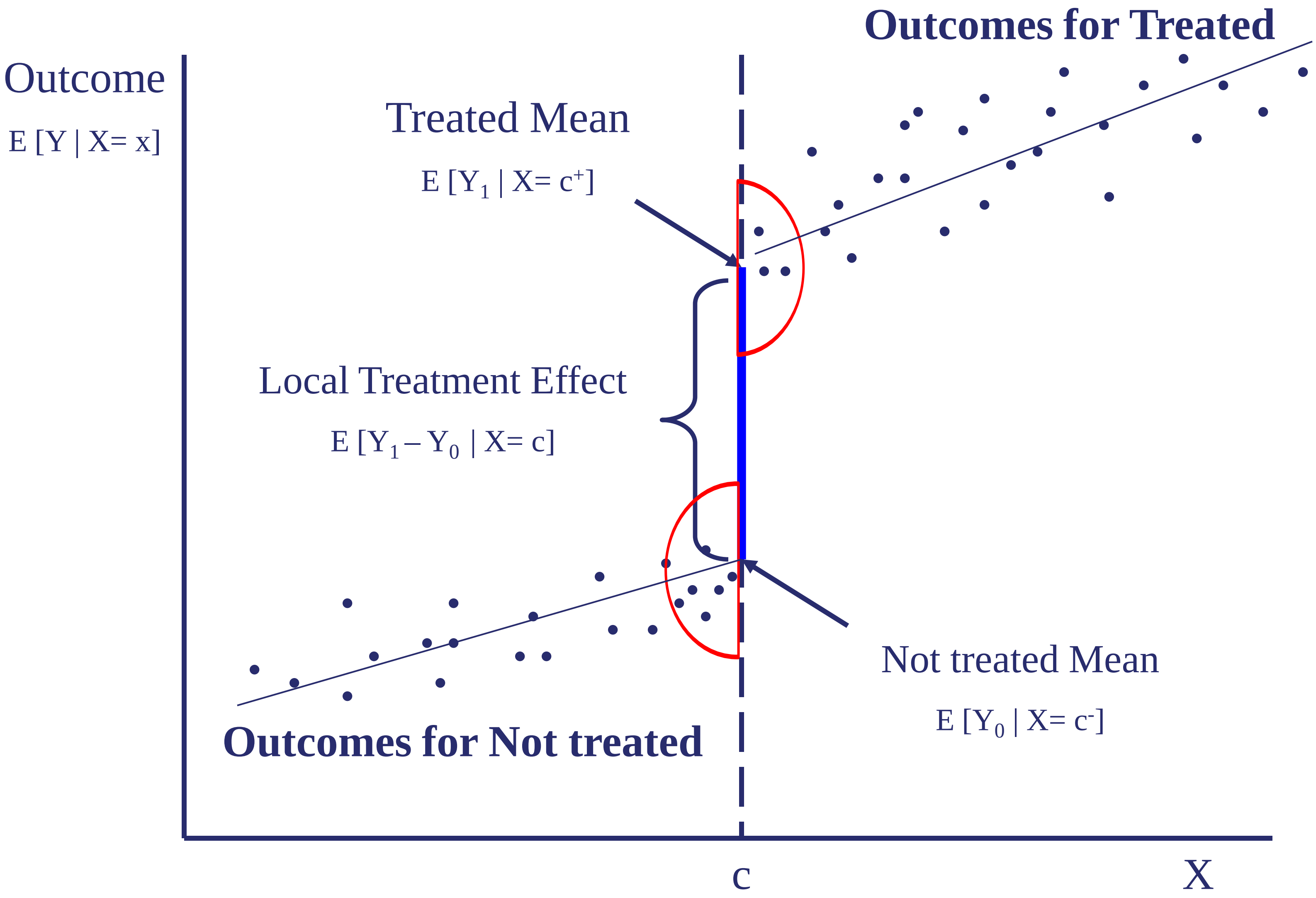
$$\lim_{d \to 0}\left\{E(Y_1 \mid X = c+d) - E(Y_0 \mid X = c-d)\right\}$$

$$= \tau + \lim_{d \to 0}\left\{E(\varepsilon \mid X = c+d) - E(\varepsilon \mid X = c-d)\right\} = \tau$$

Sharp RDD

- What does the assumption the E(□|X) is continuous (at c) mean?
  – Observations are randomly distributed at the cut point.
- With RDD the idea is that instead of using all observations to get an estimate, only use observations “near” the cut point.

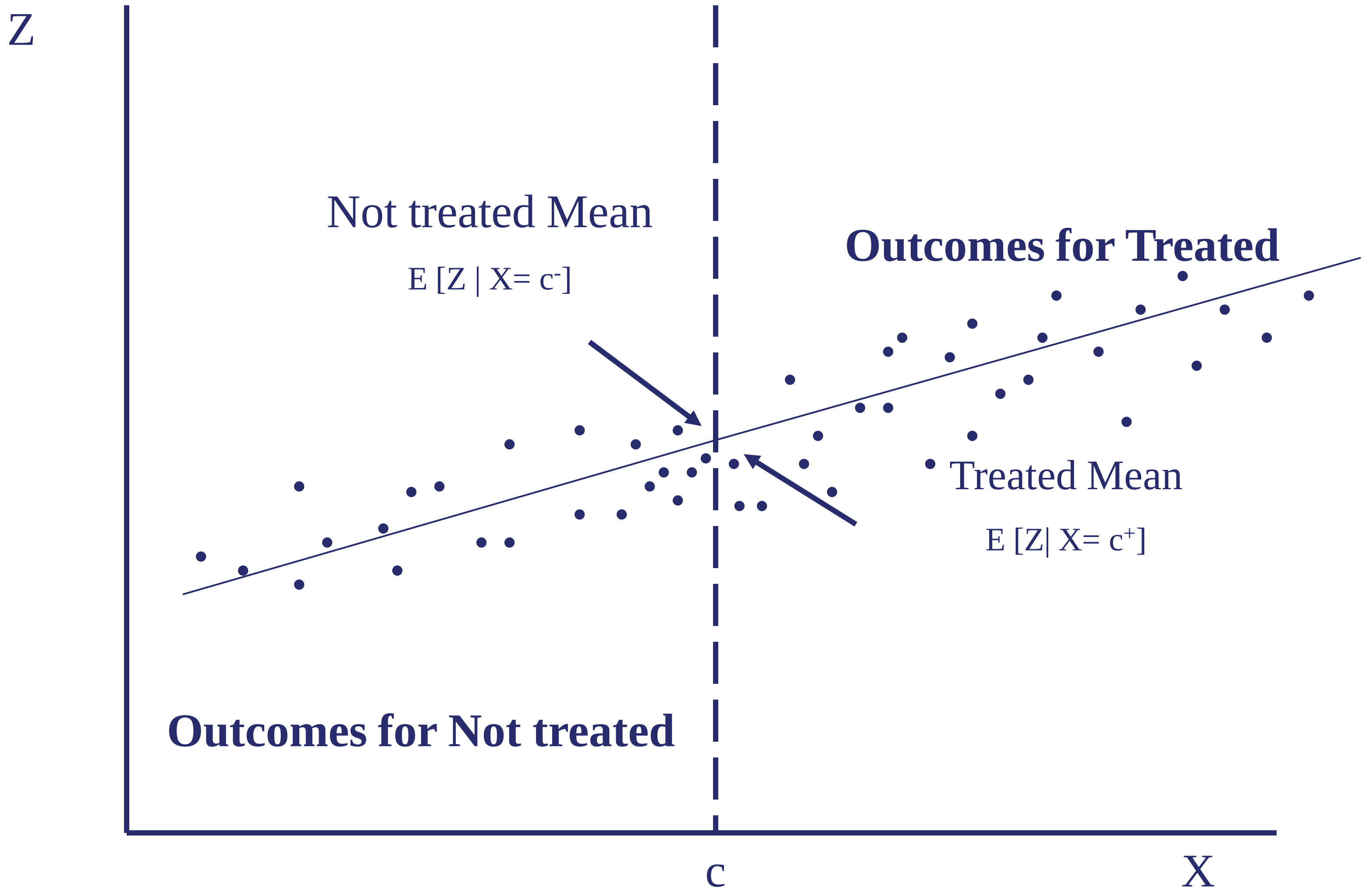
The RD Estimation Strategy



Sharp RDD

- Just like with randomized experiments it is important to check that randomization was done correctly, with RDD need to check that observations are randomly distributed around cut-point.
- Let ***Z*** denote a vector of observable pre-determined variables. Then the *E(**Z**)* should be the same on either side of the cut point.

Checking randomization around Cut-Point



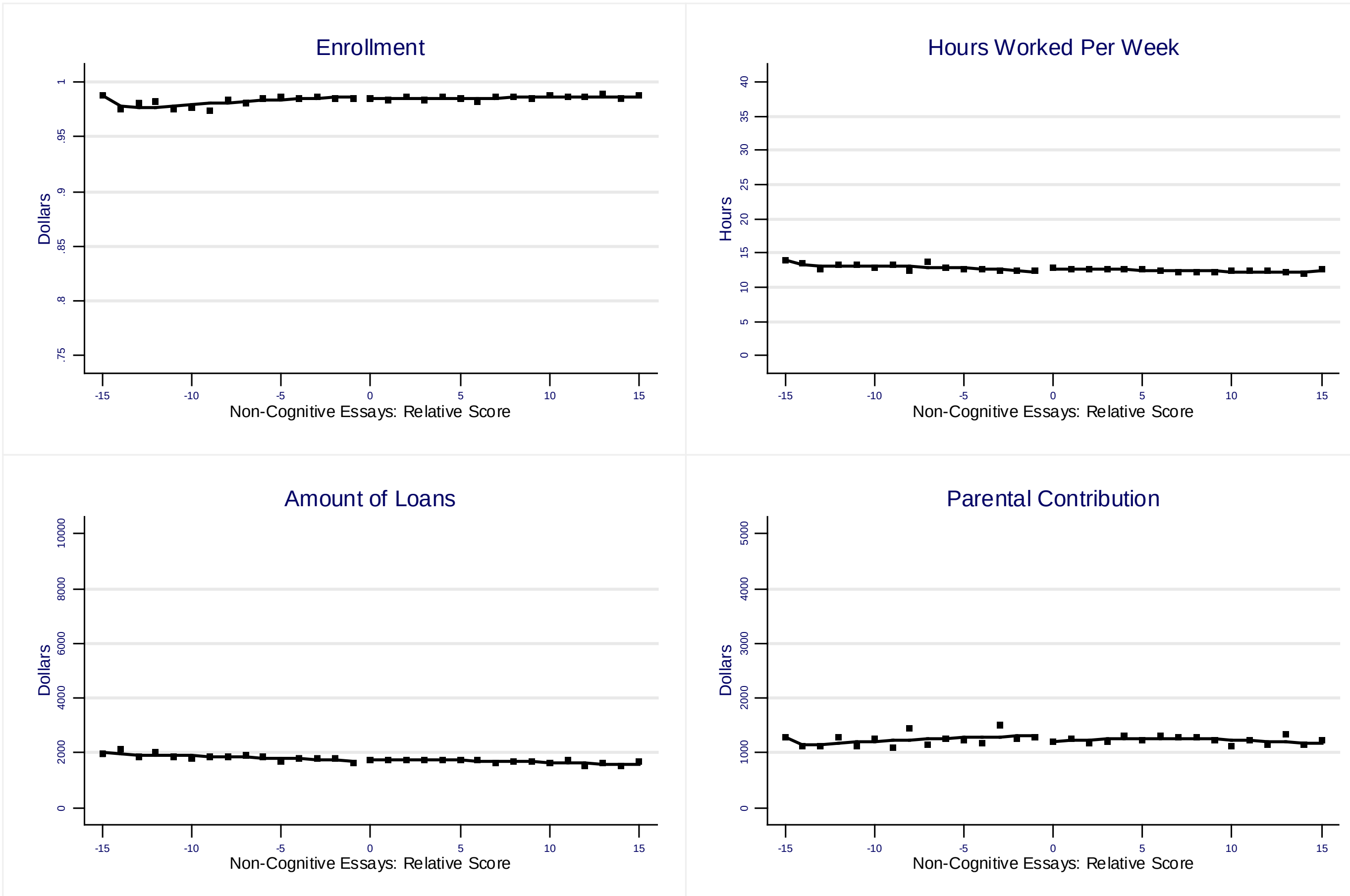
Checking randomization around Cut-Point

- There should be no “jump” in the mean of Z at cut-point for any Z or any linear combinations of the Z’s.
- One potential linear combination of relevance is to estimate a regression model of an outcome variable on the Z’s and compute the predicted value of the outcome variable for the different levels of the assignment variable.

A Legacy of Leaders Making a Difference
cshpe
1957 - 2007

Checking randomization around Cut-Point

Figure 5
Predicted Outcomes in Baseline Survey by Relative Non-Cognitive Score



Source: Gates Millennium Scholar Surveys: Cohort II & III.

No Discontinuity in Distribution of X

- Some may think that another requirement of RDD is that individuals must not be able have any control over assignment variable (X) near the cut-point.
  - For example individuals don’t know what c is.
- However, Lee (2008) showed that all you need is imperfect ability to control assignment variable.

No Discontinuity in Distribution of X

- This occurs as long as the distribution of assignment variable X is continuous at cut-point.
- So need to check for whether or not there are jumps in the probability distribution function of X at the cut point (McCrary, 2008).

No Discontinuity in Distribution of X

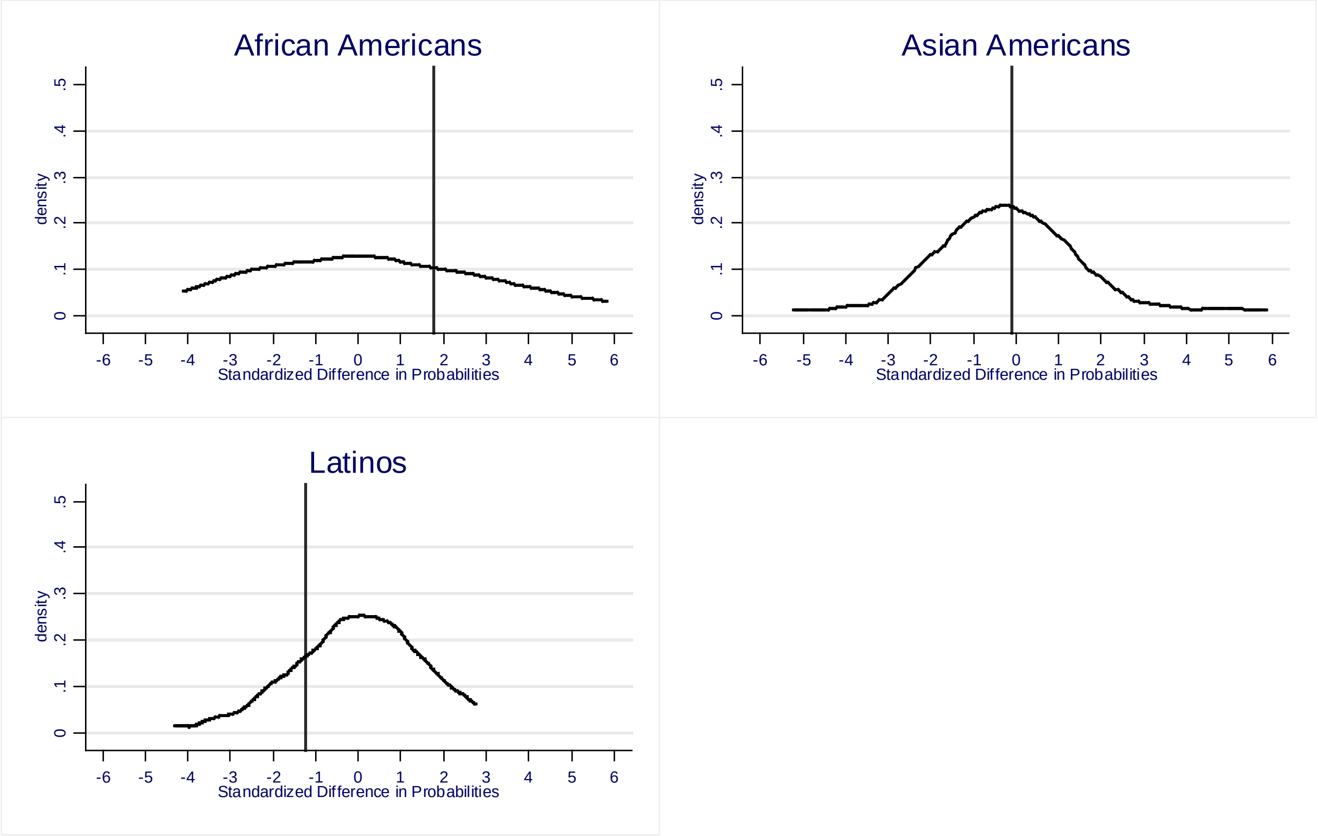
- McCrary (2008) proposes a simple two-step procedure for testing whether there is a discontinuity in the density of the assignment variable.
  - Step 1: The assignment variable is partitioned into equally spaced bins and frequencies are computed within those bins.
  - Step 2: Treats the frequency counts as the dependent variable to check for jumps at cut.

No Discontinuity in Distribution of X

- In our GMS paper we have a discrete assignment variable. So, we do something slightly different from McCrary.
  - With discrete data, will always get jumps in estimated proportions of values of X as move from x_t to x_{t+1}.
  - Question is whether the change in estimated difference in proportions, relative to the standard error of difference estimate, is large compared to other values of x_t when $x_t = c - 1$.

No Discontinuity in Distribution of X

Figure 3
Smoothed Density Estimates of Standardized Difference in Probabilities: Cohort II



Source: Gates Millennium Scholar Surveys: Cohort II. The density function estimates are based on the sample of 2340 applicants who were asked to complete the survey and are weighted to reflect the population of GMS applicants. Vertical lines indicate the standardized change between cut-point+1 and cut-point

Fuzzy RDD

- In the sharp design Pr(D=1) goes from 0 to 1 as the X crosses the cut-point c.
- **All you really need is a discontinuous jump in Pr(D=1)**:

$$\lim_{d \to 0} \{E(D \mid X = c + d) - E(D \mid X = c - d)\} > 0$$

- And

$$\tau = \frac{\lim_{d \to 0} \{E(Y \mid X = c + d) - E(Y \mid X = c - d)\}}{\lim_{d \to 0} \{E(D \mid X = c + d) - E(D \mid X = c - d)\}}$$

Fuzzy RDD

- Well known Wald formulation of treatment effect in an IV setting.
- Numerator is an estimate of the intent to treat.

RDD Estimation Strategy

- Parametric
  - Model $E(\varepsilon|x)$ as a polynomial function of x
    - $E(\varepsilon|x) = \beta_0 + \beta_1 x + \beta_2 x_2 + \beta_2 x^3$
    - $E(y|x) = \alpha E(D|x) + \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$
    - Instrument for treatment using $I(X > c)$.
    - **Need to determine what order of polynomial to use.**
    - Also may want to **restrict the sample to an interval around cut-point** to limit influence of points far away from cut-point on estimates.

RDD Estimation Strategy

- **Graphical Presentation**
- For some bandwidth h and for some number of bins K_0 and K_1 to the left and right of the cut-point, respectively, want to construct bins $(b_k, b_{k+1}]$ of length h and compute average value of Y in bin.

$$\bar{Y}_k = \frac{\sum_{i=1}^{N} Y_i I\{b_k < X_i \le b_{k+1}\}}{\sum_{i=1}^{N} I\{b_k < X_i \le b_{k+1}\}} = \frac{\sum_{i=1}^{N} Y_i I\{b_k < X_i \le b_{k+1}\}}{N_k}$$

RDD Estimation Strategy

- What bin width should you use?
  - One choice is to choose width that minimizes cross-validation function:

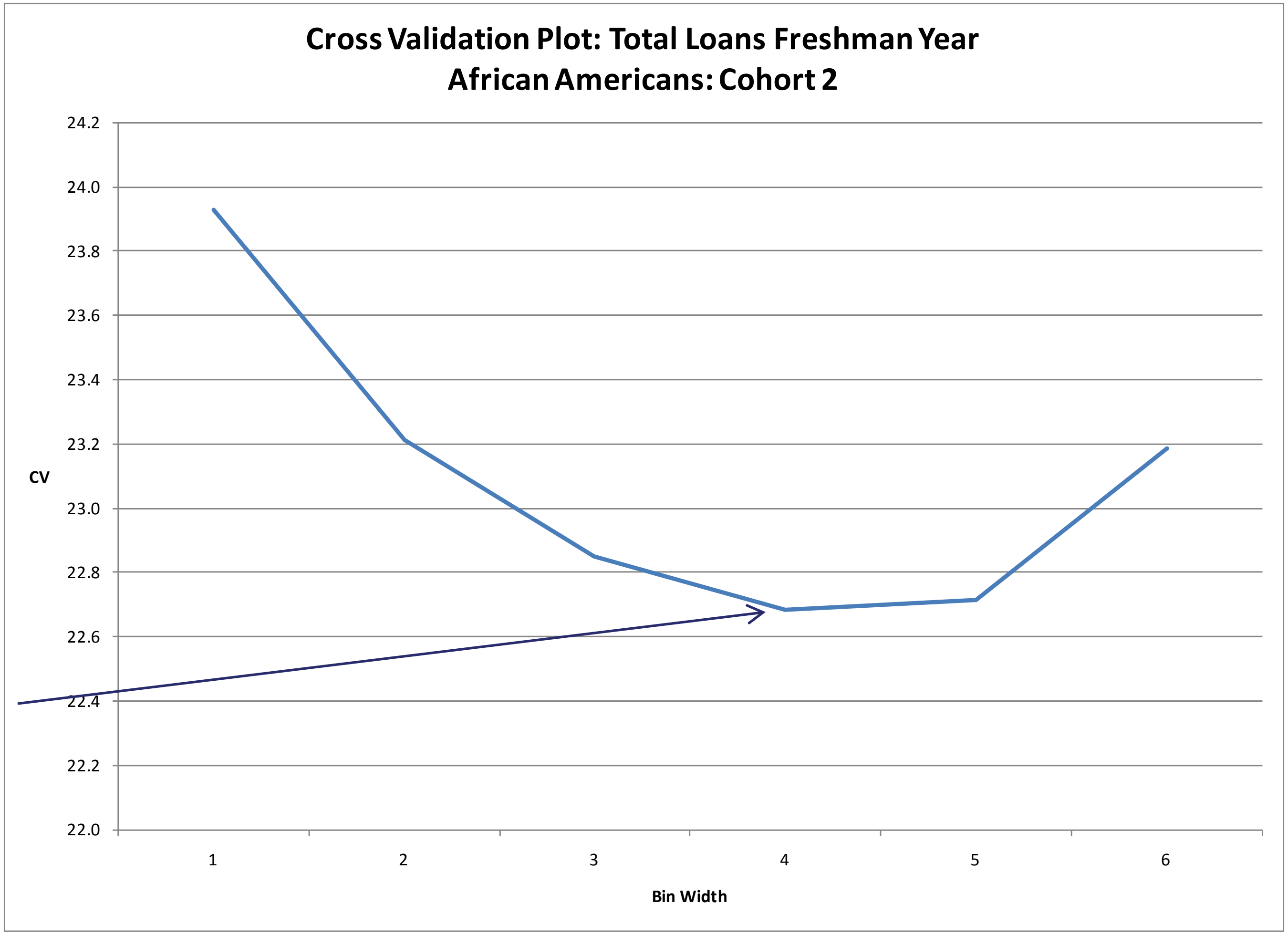
$$CV_Y(h) = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2$$

  - Where

$$\hat{Y}_i = \frac{1}{N_k - 1}\sum_{j \neq i} Y_j I(b_k < X_j \leq b_{k+1})$$

and $X_i \square (b_k, b_{k+1}]$.

RDD Estimation Strategy



Minimized at bin width of 4

RDD Estimation Strategy

- **<u>Non-parametric estimation</u>**
  - **Sharp design**
  - $Y = m(X) + \square D + \square$
  - Use local polynomial (linear) regression to estimate regression line “just below” and “just above” the cut point.
    - Estimate model first only using data above the cut point and then using data only below cut point.

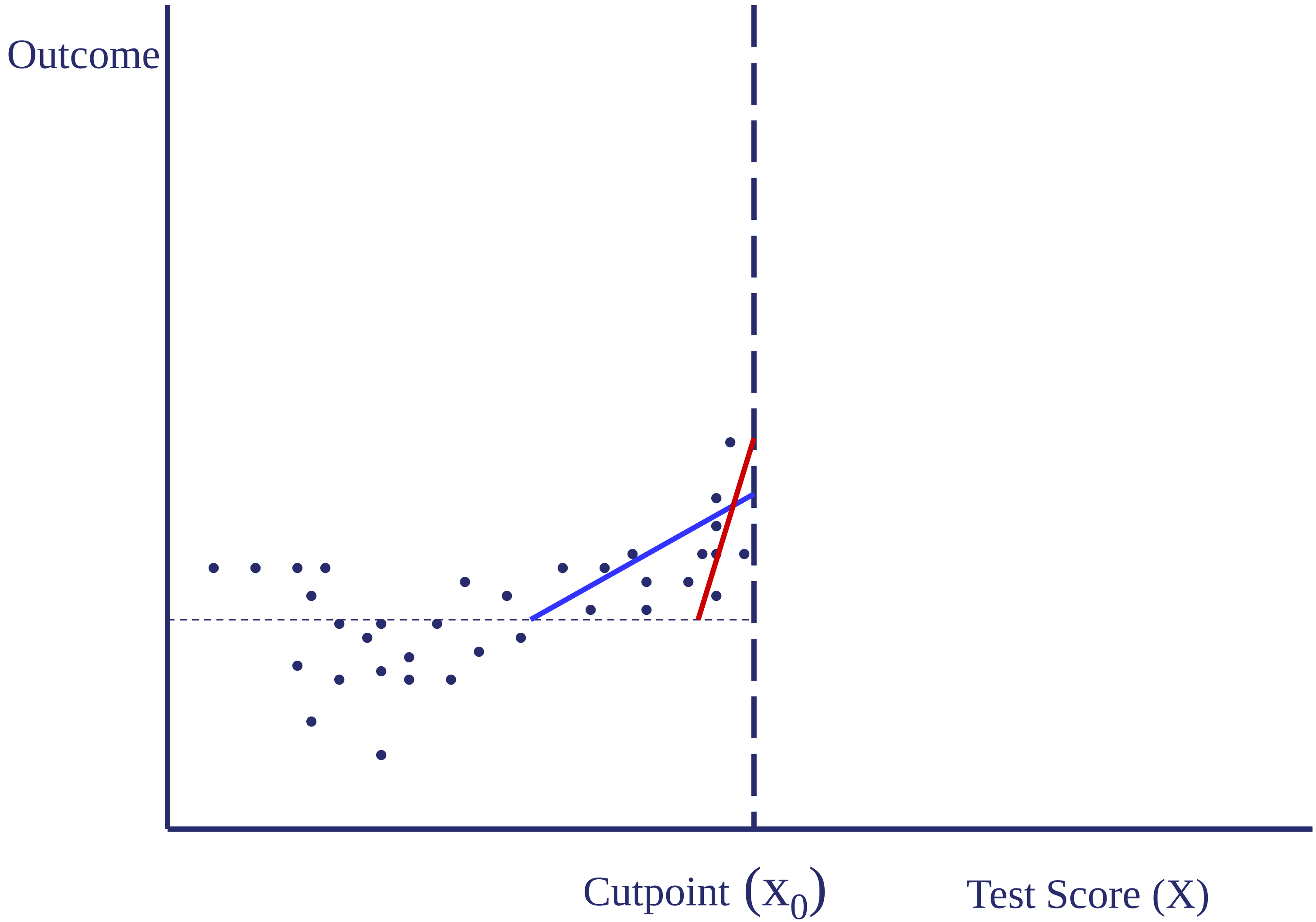
$\hat{m}(X)$ estimate above cut

$\tilde{m}(X)$ estimate below cut

RDD Estimation Strategy

- Then $\hat{\tau} = \lim_{X \downarrow c} \hat{m}(x) - \lim_{X \uparrow c} \tilde{m}(x)$
- How to estimate m(X)? Use Local polynomial regression.
- Local polynomial regression is a series of weighted regressions. Use kernel density function (K(t)) to determine weights (e.g. Gaussian distribution) along with a bandwidth (h).
- How do we choose bandwidth h?

Optimal Bandwidth Selection



RDD Estimation Strategy

- Optimal Bandwidth at a point x_0 balances Variance and Bias so as to minimize mean squared error.

$$\text{Bias} = K_1 m''(x_0) h^2 \qquad \text{Var} = K_2 \frac{\sigma^2(x_0)}{f(x_0)nh}$$

$m''(x_0)$ measure of the curvature at x_0

Choose bandwidth h to minimize Bias² + Var :

$$h_{opt}(x_0) = C\left[\frac{\sigma^2(x_0)}{\{m''(x_0)\}^2 f(x_0)}\right]^{1/5} n^{-1/5}$$

where C depends on kernel choice.

RDD Estimation Strategy

- <u>Estimation strategy for optimal bandwidth</u>
- Separately for those below cut-point (x_0)
  - Estimate 4th order polynomial regression of dependent variable on non-cognitive test score and compute

$$\lim_{x \uparrow x_0} \tilde{m}''(x) \equiv \tilde{m}''(x_-) \quad \text{and} \quad \lim_{x \uparrow x_0} \tilde{\sigma}(x) \equiv \tilde{\sigma}(x_-)$$

  - **Compute rule of thumb bandwidth** by minimizing mean squared error.
  - **Estimate local cubic polynomial regression** using rule of thumb bandwidth.

RD Estimation Strategy

- Compute $\hat{m}''(x\_)$ and $\hat{\sigma}(x\_)$ from local cubic polynomial regression and **estimate optimal bandwidth**.
- **Estimate local linear regression** using optimal bandwidth.

- **Repeat for those above the cut-point.**
- Data-driven or plug-in bandwidth.