Regression Discontinuity Designs

Example – The Effect of MADS on Earnings

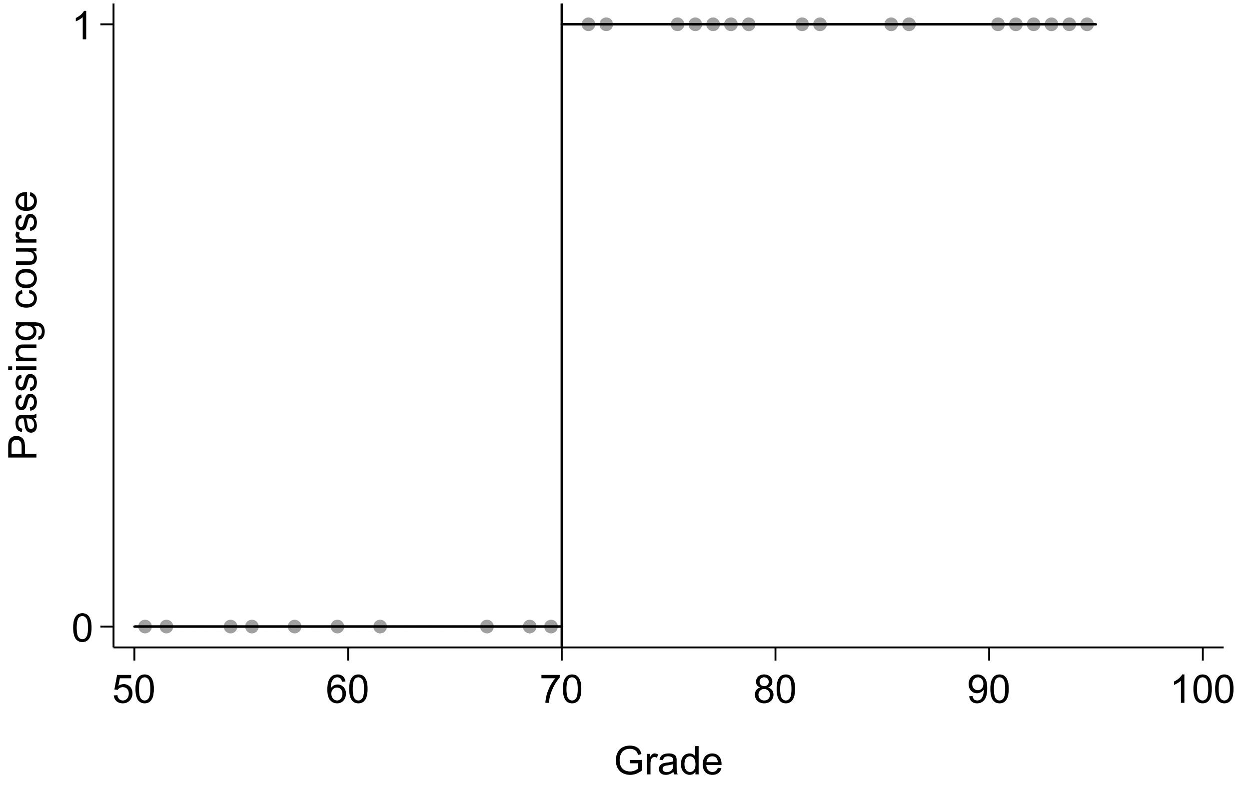
Imagine you want to estimate the effect of passing the MADS causal inference course on earned income

- Randomly assigning some students to the “passing” treatment and failing others is unethical
- Controlled regression might be biased because of unobserved confounders
- Passing cutoff creates a natural experiment

Emily Glassberg Sands: https://medium.com/teconomics-blog/5-tricks-when-ab-testing-is-off-the-table-f2637e9f15a5

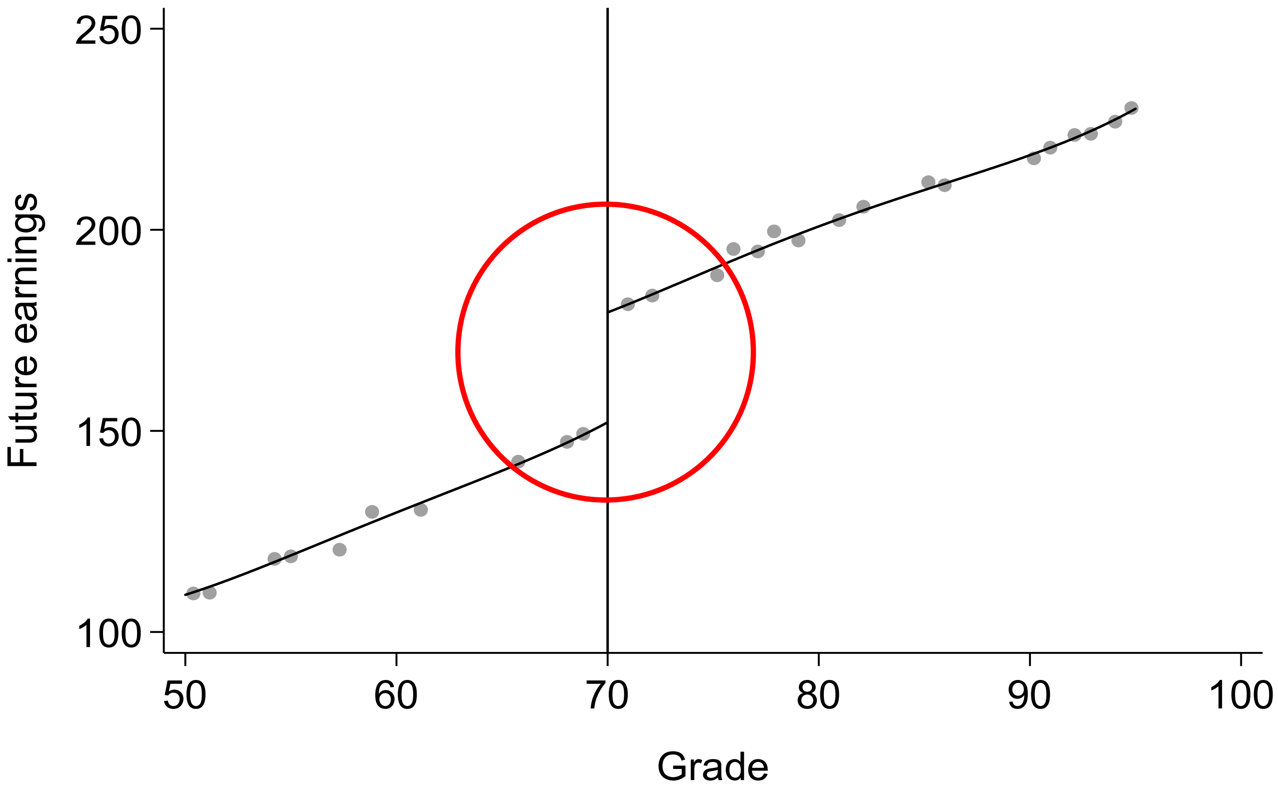
Example – The Effect of MADS on Earnings

Discontinuity in passing the course:



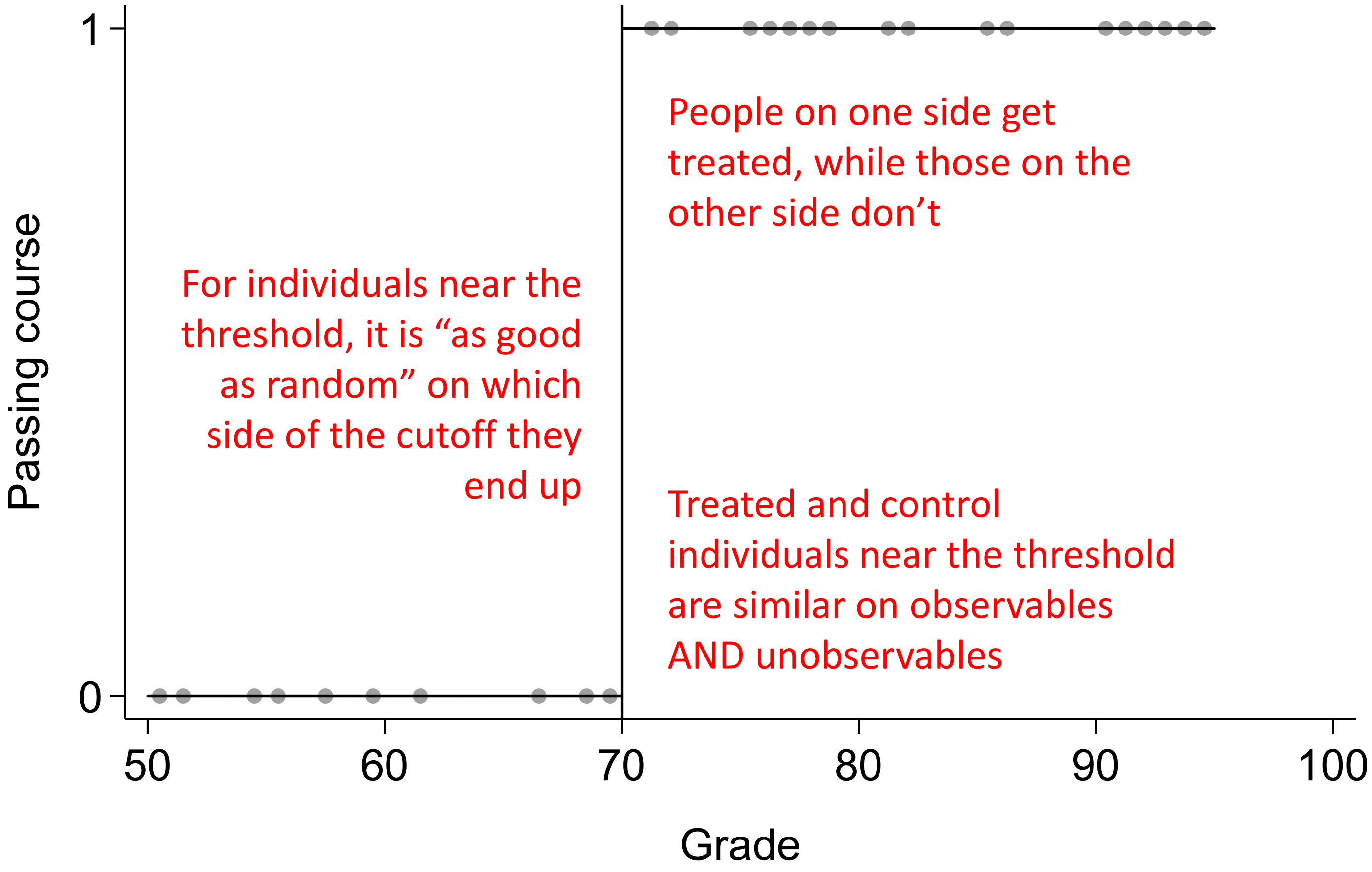
Example – The Effect of MADS on Earnings

Causal effect of passing on earned income:



Regression Discontinuity Designs

Regression discontinuity (RD) designs exploit “arbitrary” cutoffs embedded in rules



Basic RD Setup

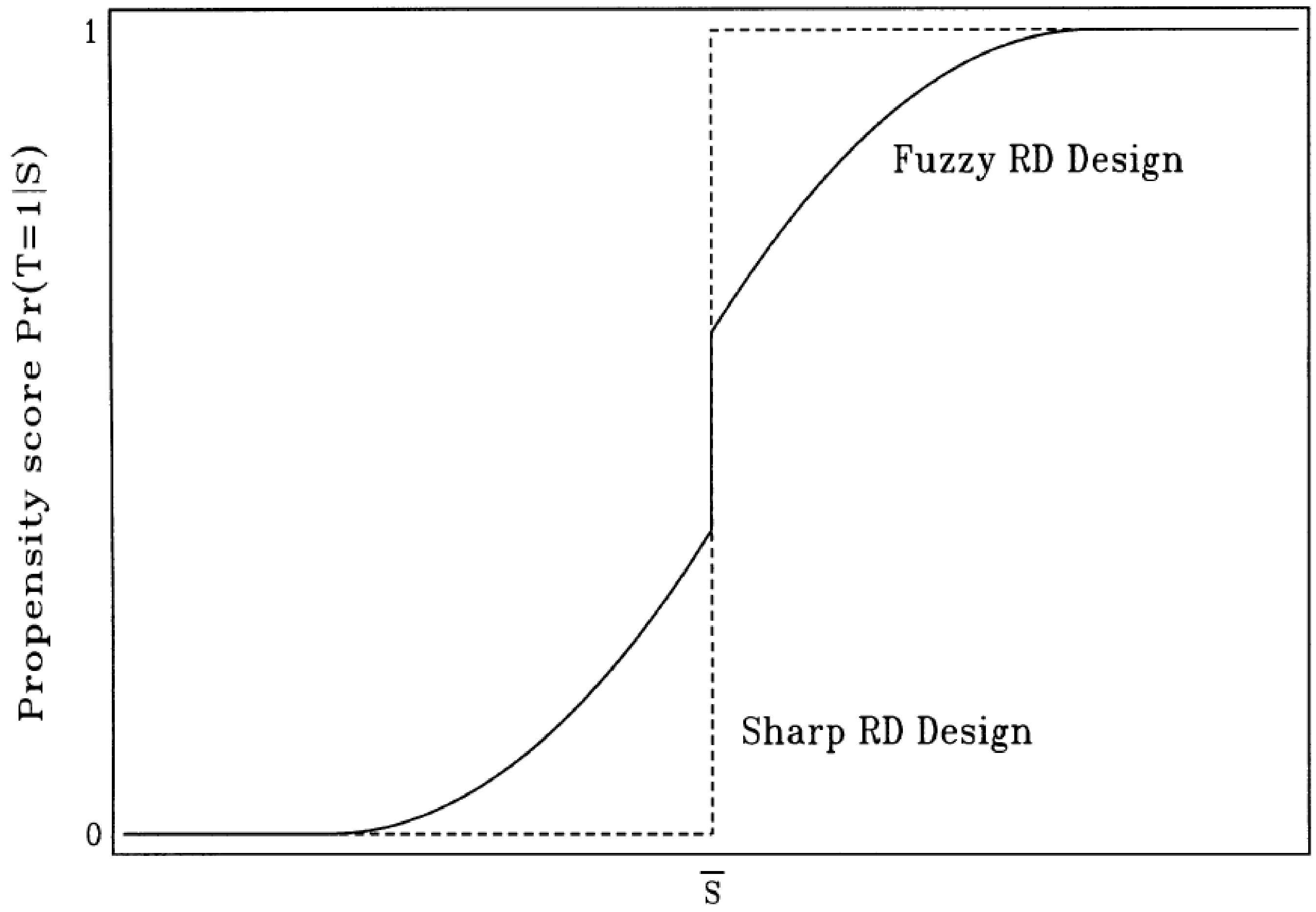
- X_i is the **running variable**

- Treatment assignment is determined by a cutoff in X_i

- Relationship between potential outcomes and running variable has to be “smooth” at the cutoff

- Changes in the outcome around the threshold can be interpreted as causal effects

Sharp vs. Fuzzy RD Designs



van der Klaauw (2002)

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Credits:
Alain Cohn
Assistant Professor of Information

Sharp Regression Discontinuity Designs

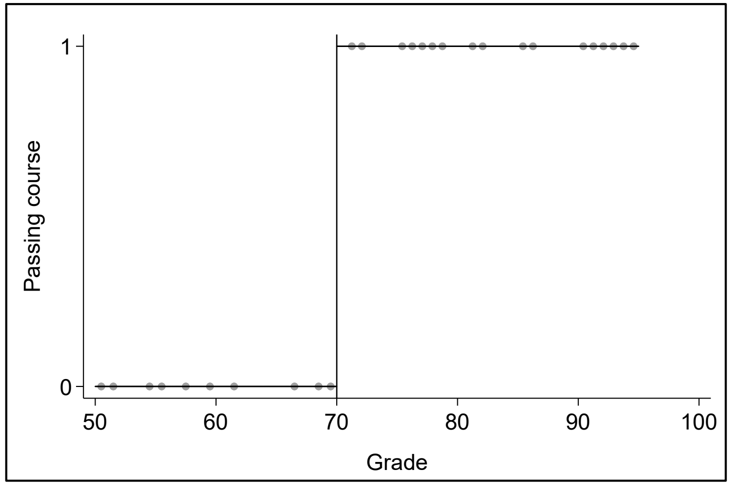
Sharp RD

Treatment assignment is a *deterministic* function of the running variable X_i and the cutoff c:

$$D_i = \begin{cases} 1 & \text{if} \quad X_i \geq c \\ 0 & \text{if} \quad X_i < c \end{cases}$$

Example: Passing course on earned income

- If grade is 70 or above → Pass
- If grade is below 70 → Not pass



RD and Potential Outcomes

Assuming constant effects and linearity in X_i:

$$Y_i^0 = \alpha + \beta X_i$$
$$Y_i^1 = Y_i^0 + \tau$$

Using the switching equation $Y_i = Y_i^0 + \left(Y_i^1 - Y_i^0\right)D_i$, we get:

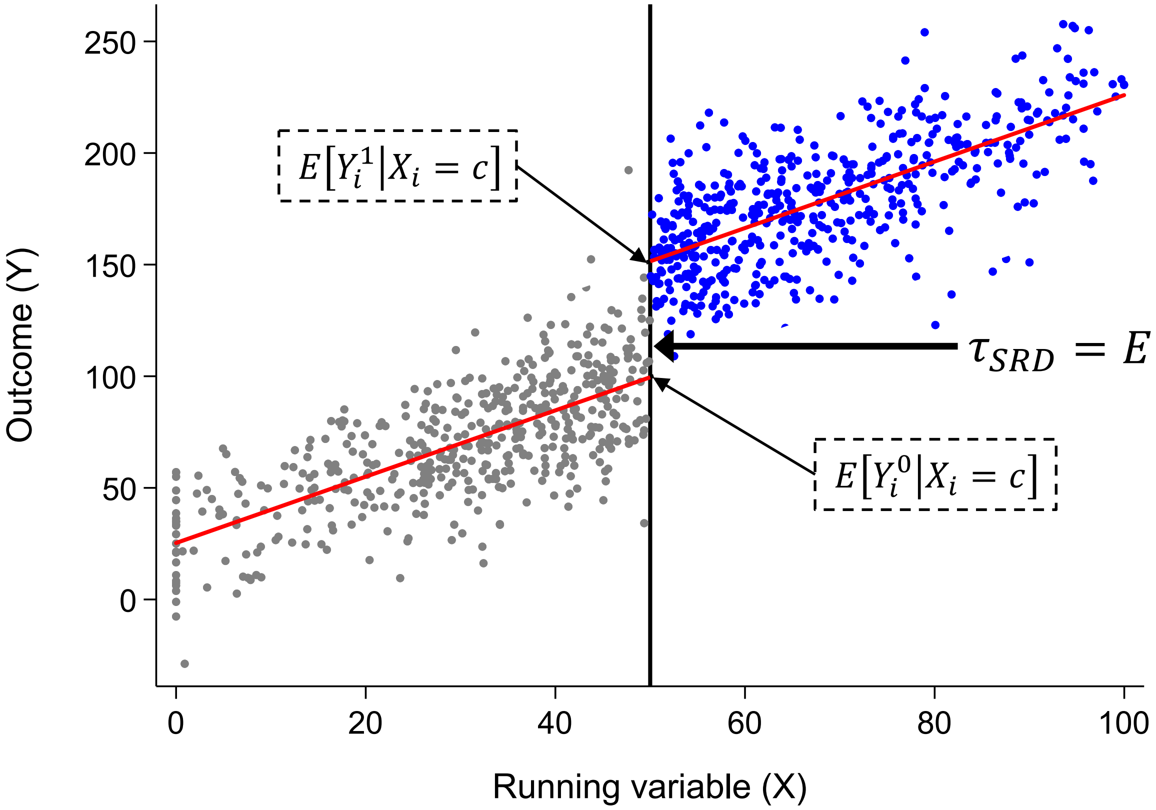
$$Y_i = \alpha + \beta X_i + \tau D_i + \epsilon_i$$

Sharp RD designs estimate the ATE at the threshold:

$$\tau_{SRD} = E[Y_i^1 - Y_i^0 | X_i = c]$$

Extrapolation

We don't observe $E[Y_i^0|X_i = c]$, so we have to extrapolate from $E[Y_i^0|X_i = c - \epsilon]$ and vice versa:



Dashed lines are extrapolations

$\tau_{SRD} = E\left[Y_i^1|X_i = c\right] - E[Y_i^0|X_i = c]$

Continuity Assumption

The conditional expectation functions of the potential outcomes

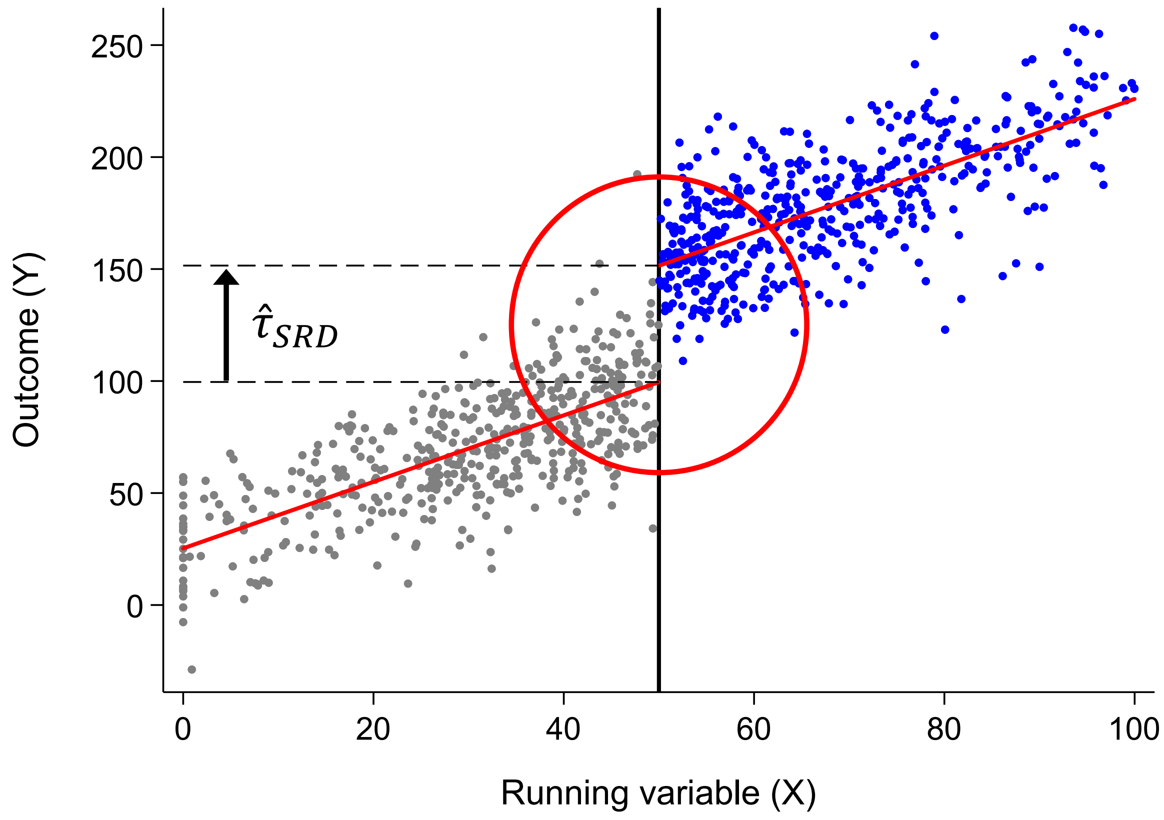
$$E[Y_i^0|X_i = x] \text{ and } E[Y_i^1|X_i = x]$$

are continuous (smooth) in x.

- *Imperfect control over X_i*: individuals cannot control whether they are just above (or below) the cutoff
- *No confounding discontinuities*: being just above (or below) the cutoff should not influence other covariates

“Local” Average Treatment Effect

Sharp RD treatment effect is a narrow or “local” ATE around the cutoff:



Scott Cunningham: “Causal Inference: The Mixtape”

Minimum Legal Drinking Age

In 2008, a number of college presidents have endorsed the Amethyst Initiative, which calls for a reexamination of the minimum legal drinking age (MLDA)

- The age-21 limit in the U.S. is higher than in most Western countries
- A key argument of the Amethyst Initiative is that 18-20 year-olds would drink less if it were legal for them to drink
- Opponents say that the age-21 MLDA reduces youth access to alcohol, thereby preventing harm

Minimum Legal Drinking Age

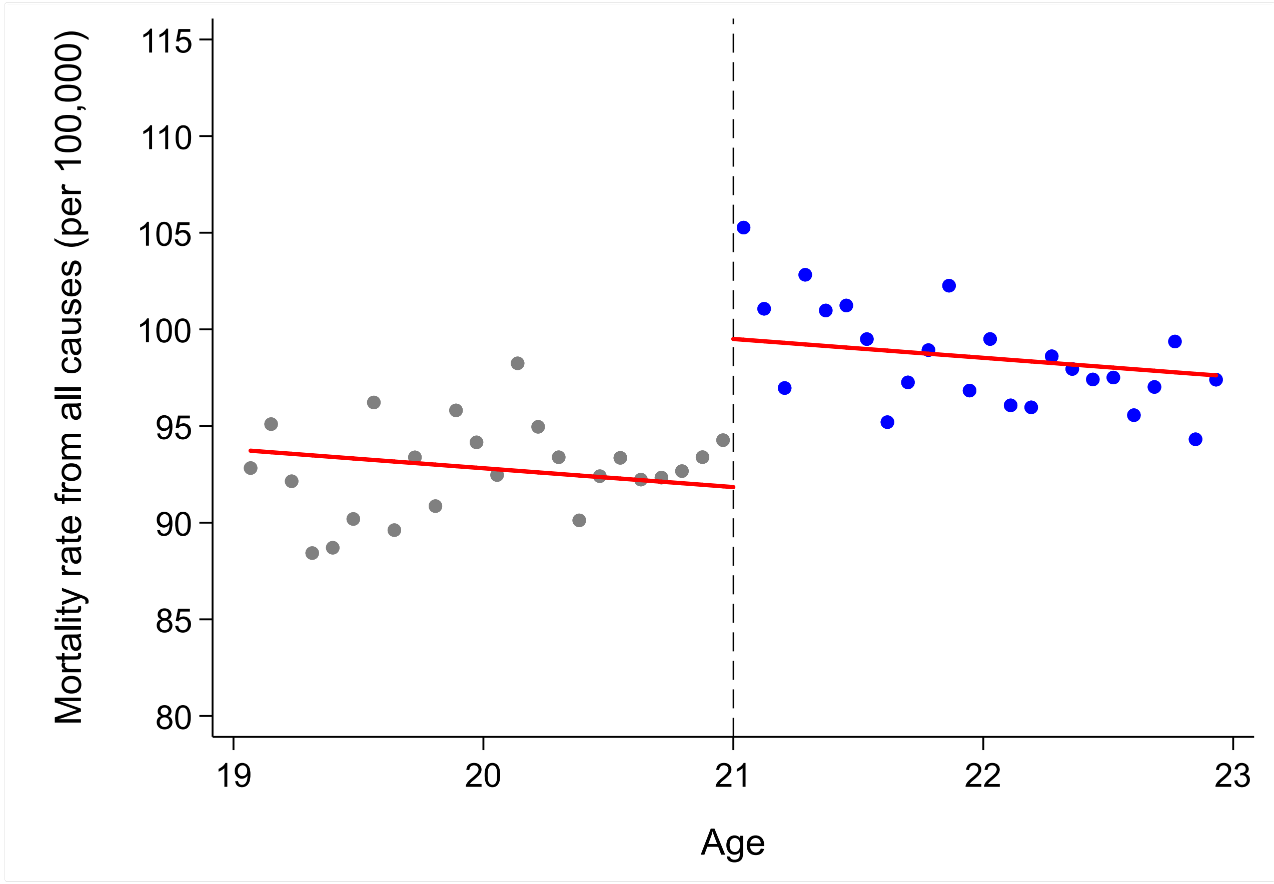
Causal question: **What is the effect of legal access to alcohol on death rates?**

- Sharp discontinuity in legal access to alcohol:
  - People just above 21 have legal access while those just below 21 don’t
  - The two groups have similar characteristics otherwise (both observed and unobserved)

- If nothing else changes abruptly at age 21, then a discrete change in death rates at age 21 can be plausibly attributed to the legal drinking age

Minimum Legal Drinking Age

Death rates as a function of age:



From *Mastering 'Metrics: The Path from Cause to Effect*. © 2015 Princeton University

Estimation in the Sharp RD

Estimate effect with a local linear regression

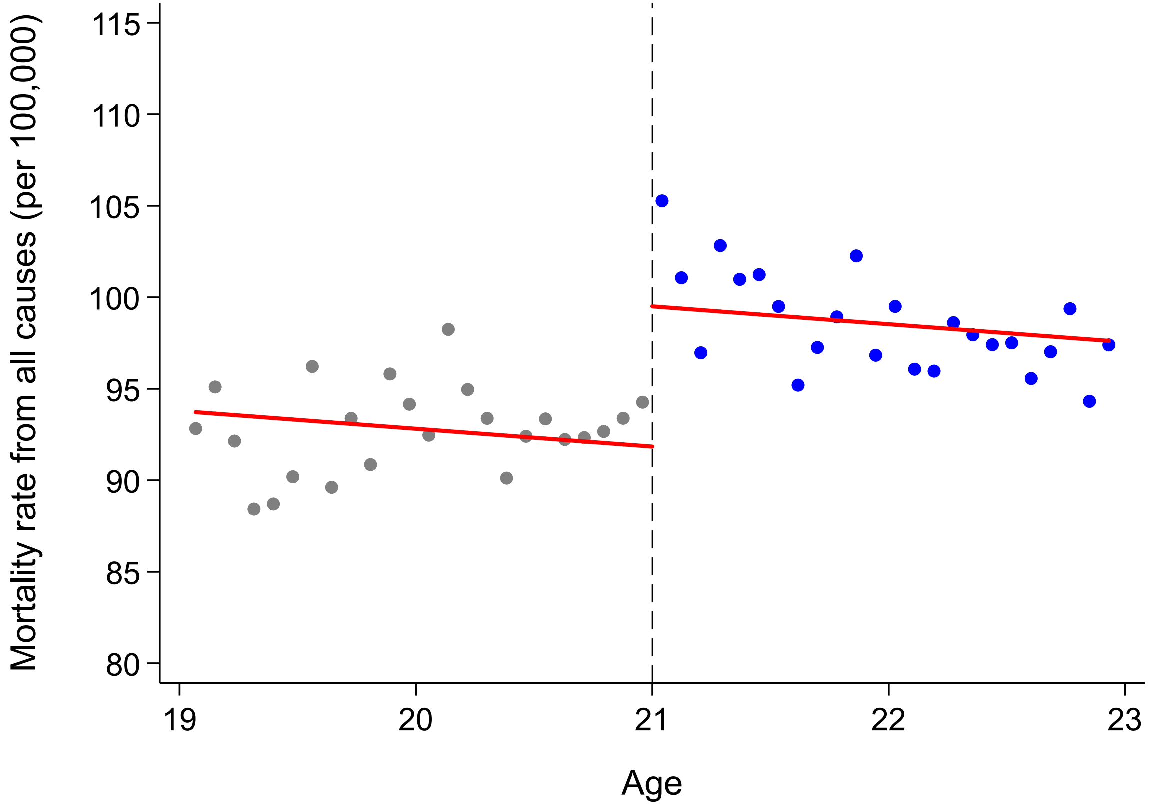
Death rates change with age (running variable)

- Not a problem if confounders change continuously

RD regression controls for smooth age-related variation in death rates:

$$Y_i = \alpha + \beta(X_i - c) + \tau D_i + \epsilon_i$$

- Unit of observation are age groups by months
- Y_i is the death rate for month i
- D_i is a dummy variable that takes a value of 1 if the age group is 21 or above (and 0 otherwise)

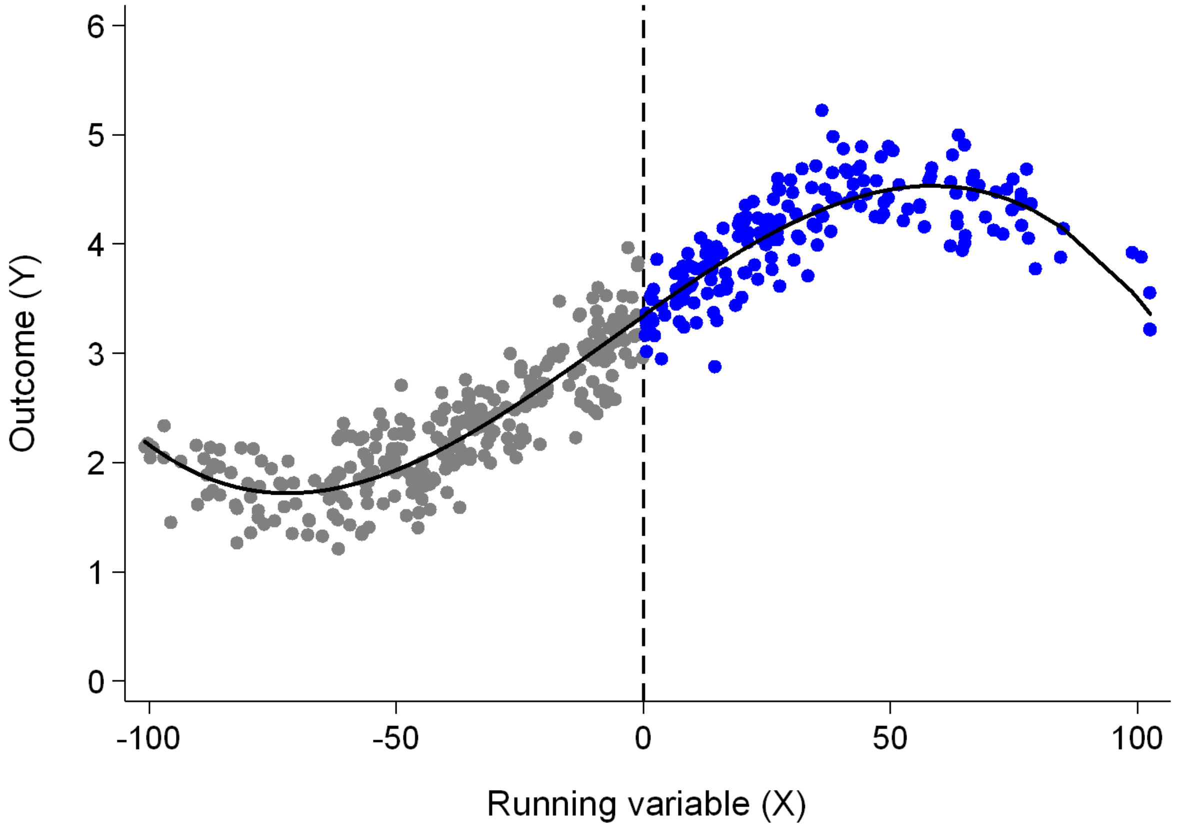


Estimation Results

```
. reg all age over21

      Source |       SS           df       MS      Number of obs   =        48
-------------+----------------------------------   F(2, 45)        =     32.99
       Model |  410.138151         2  205.069075   Prob > F        =    0.0000
    Residual |  279.682408        45  6.21516463   R-squared       =    0.5946
-------------+----------------------------------   Adj R-squared   =    0.5765
       Total |  689.820559        47  14.6770332   Root MSE        =     2.493

------------------------------------------------------------------------------
         all |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.9746843   .6324613    -1.54   0.130    -2.248527    .2991581
      over21 |   7.662709   1.440286     5.32   0.000     4.761824    10.56359
       _cons |   91.84137   .8050394   114.08   0.000     90.21994     93.4628
------------------------------------------------------------------------------
```

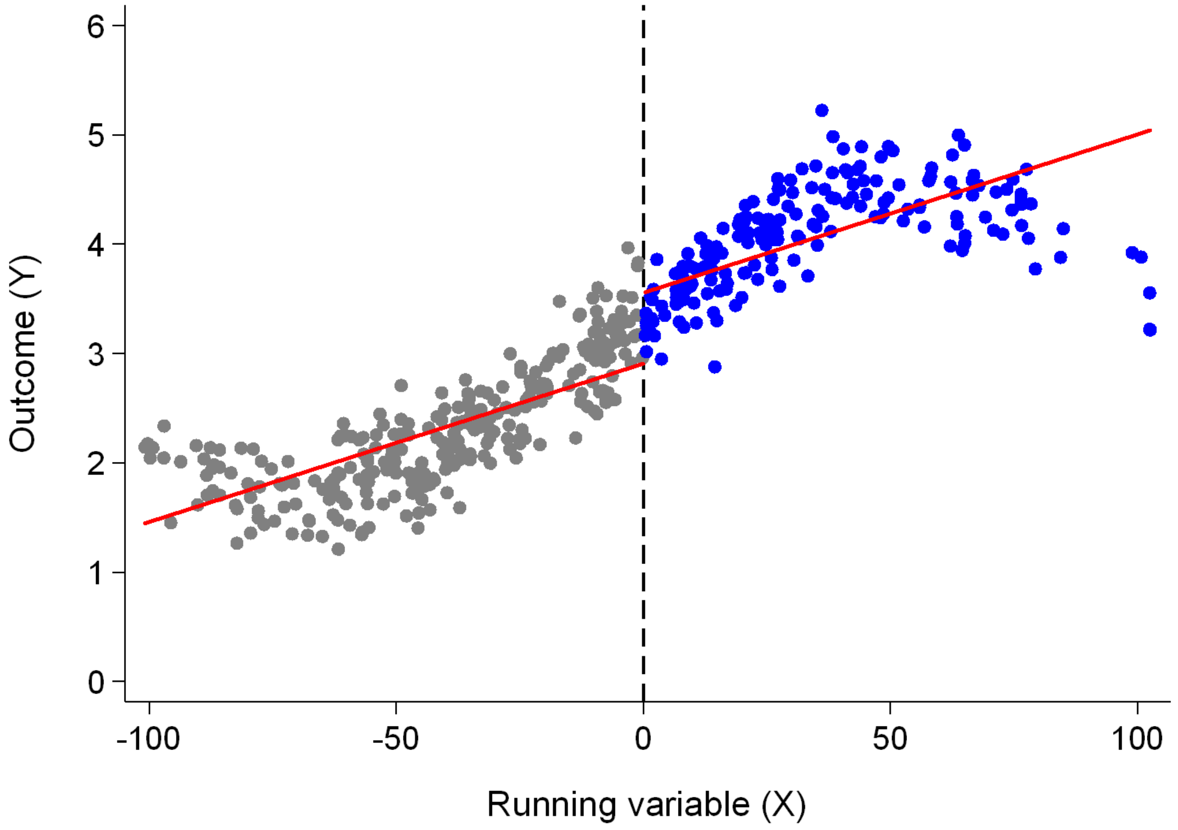
Nonlinearity Bias

What if the trend relation does not jump at the cutoff but is simply nonlinear?



Nonlinearity Bias

Using the previous regression model, we would wrongly conclude that there is a jump at the cutoff:



Modeling Nonlinearities

Add polynomials of the running variable (let $\tilde{X}_i = X_i - c$):

$$Y_i = \alpha + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \tau D_i + \epsilon_i \qquad (1)$$

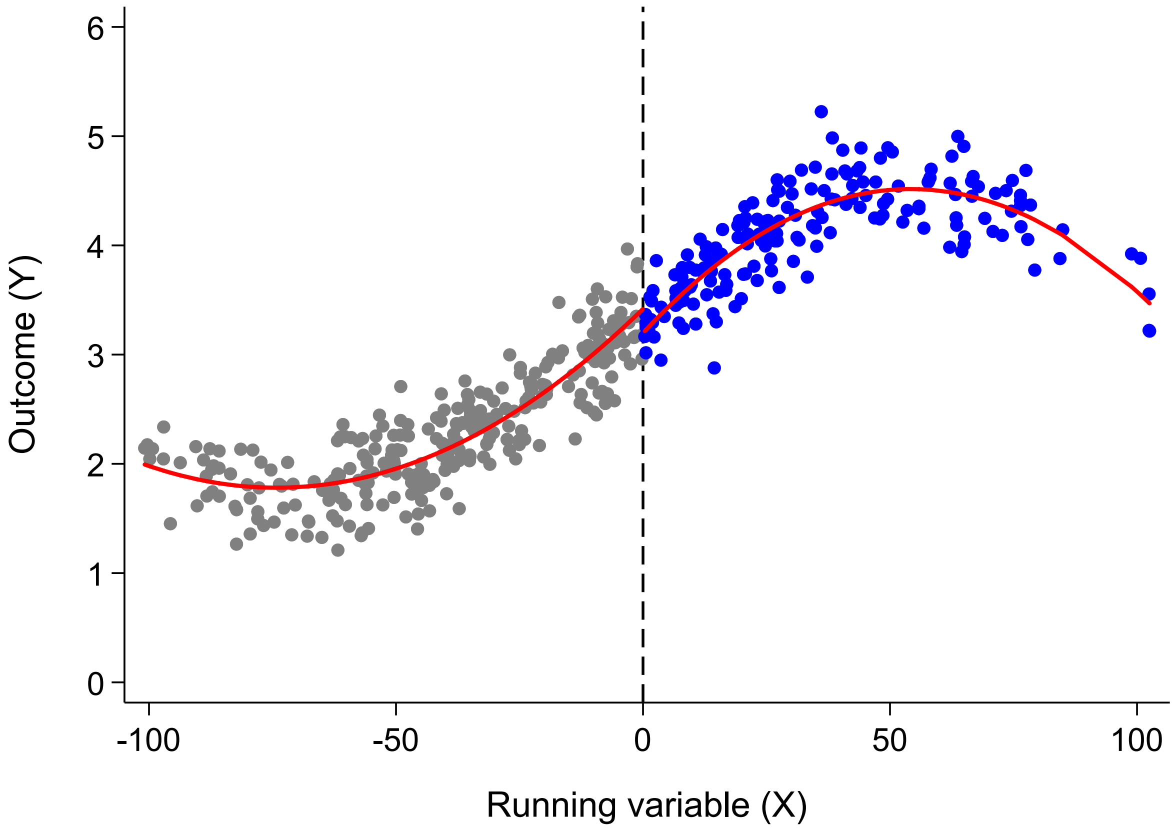
Include interaction term with D_i to allow the $\tilde{X}_i$ term to differ on both sides of the cutoff:

$$Y_i = \alpha + \beta \tilde{X}_i + \tau D_i + \gamma \tilde{X}_i D_i + \epsilon_i \quad (2)$$

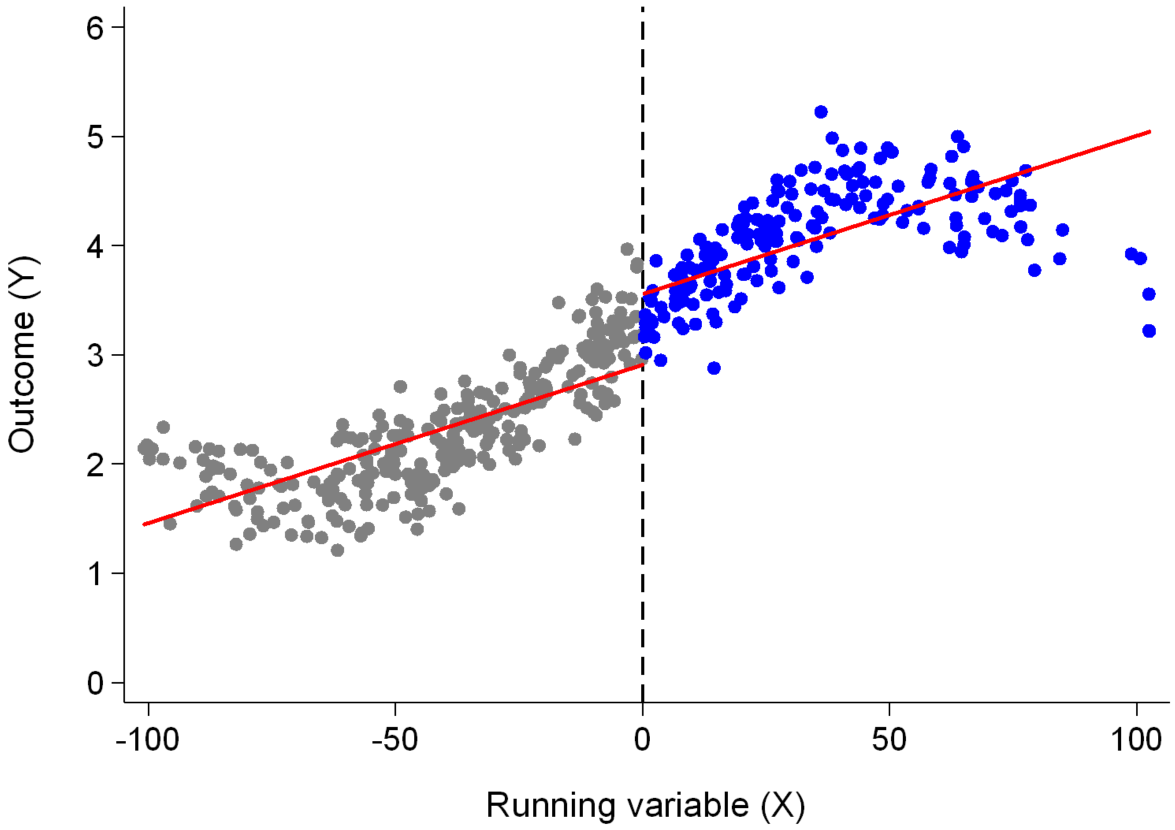
We can also combine (1) and (2)

Modeling Nonlinearities

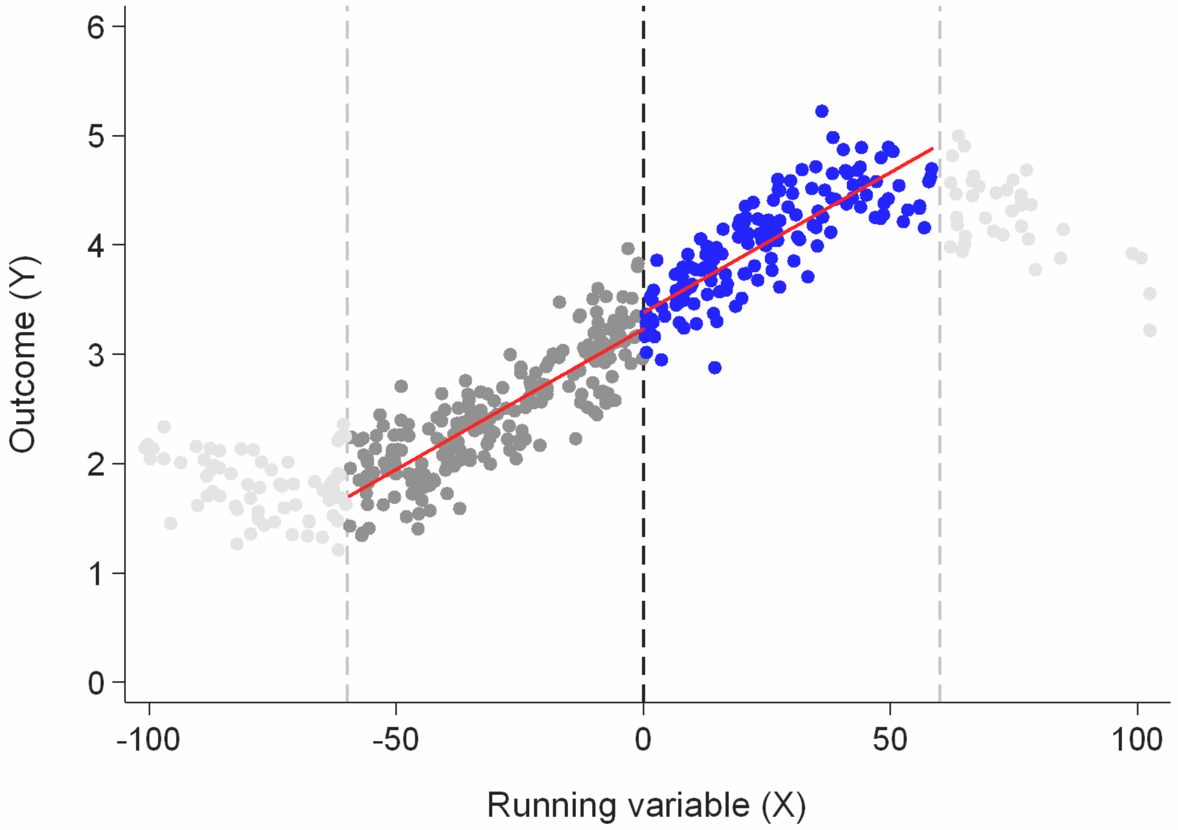
$$Y_i = \alpha + \beta\tilde{X}_i + \beta_2\tilde{X}_i^2 + \tau D_i + \gamma_1\tilde{X}_i D_i + \gamma_2\tilde{X}_i^2 D_i + \epsilon_i$$



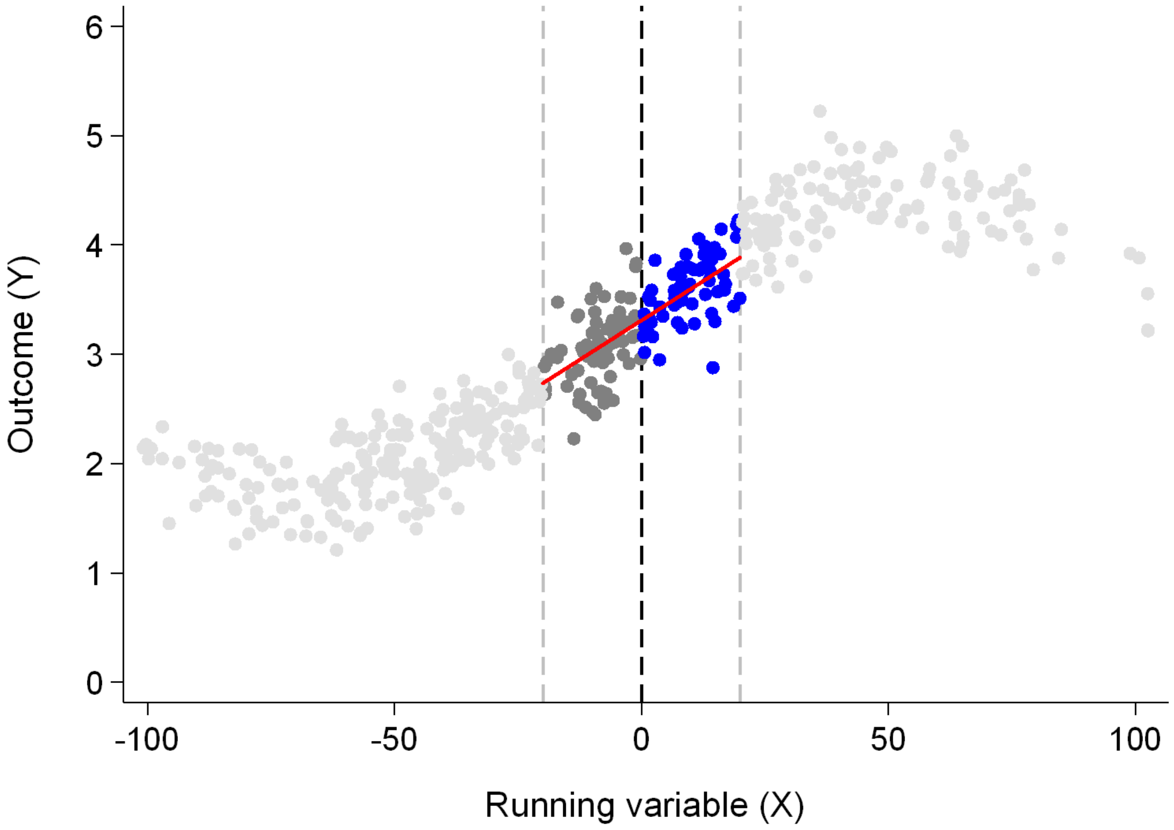
Reducing the Bandwidth



Bandwidth of 60



Bandwidth of 20



Bandwidth Selection

- Choice of bandwidth is a trade-off between *bias* and *variance*
  - Bandwidth ↑: High bias (farther from the cutoff), low variance (more data points)
  - Bandwidth ↓: Low bias (closer to the cutoff), high variance (fewer data points)

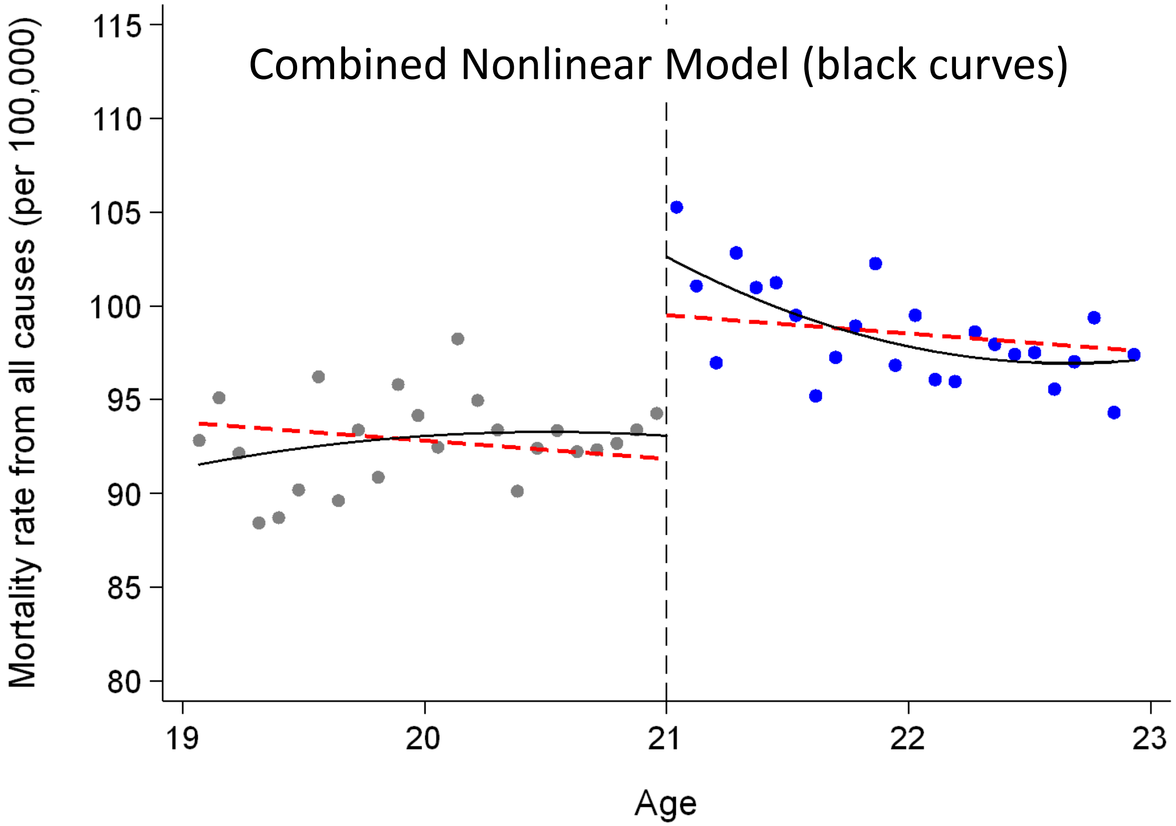
- Optimal bandwidth: Imbens and Kalyanaraman (2011)

- In practice, bandwidth choice requires a judgment call

- Are the results robust to different bandwidth choices?

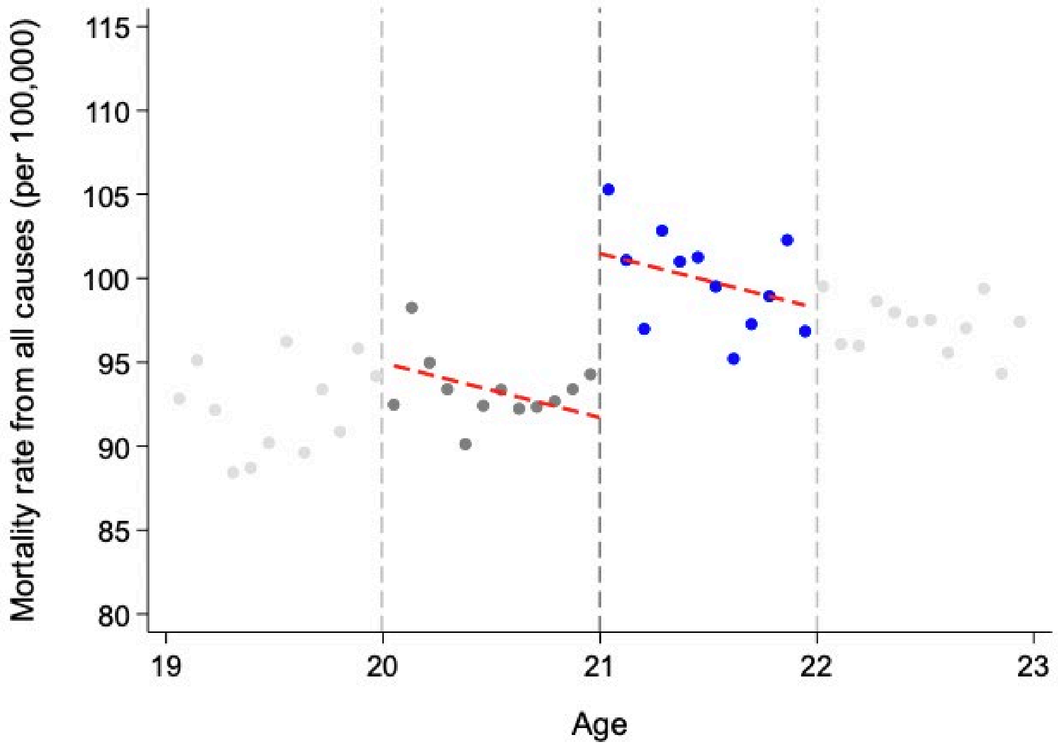
Minimum Legal Drinking Age

Comparing linear and nonlinear models:



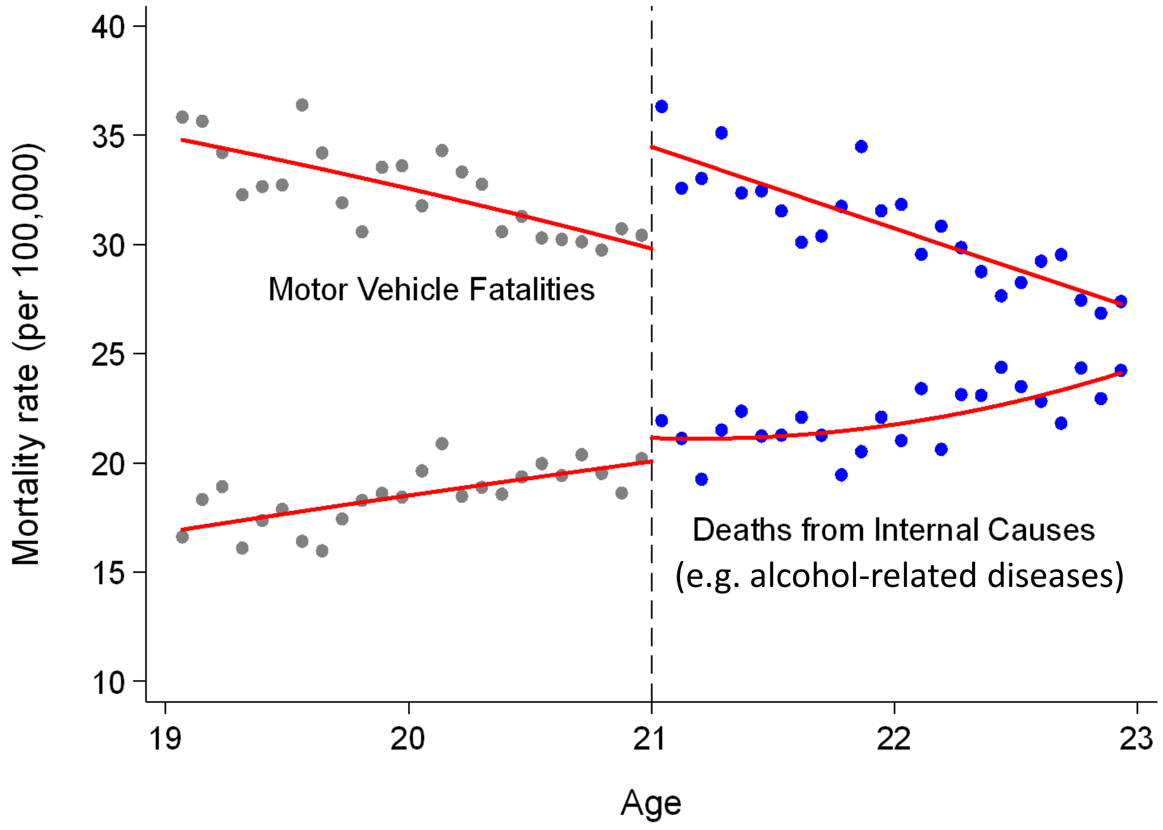
Minimum Legal Drinking Age

Check robustness by reducing bandwidth:



Placebo Test

Is the jump in death rates indeed caused by excessive drinking?



SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Credits:
Alain Cohn
Assistant Professor of Information

Fuzzy Regression Discontinuity Designs

Fuzzy RD

- Discontinuity in the *probability* of treatment at the cutoff

- But unlike with sharp RD, the probability of treatment does not go up from 0 to 1

- Example: MADS admissions test
  - If test score is above threshold → Admitted, but some do not enroll
  - If test score is below threshold → Not admitted, but some may still find a way in

- Fuzzy RD is often used when a threshold encourages treatment take-up, but does not force people to get treatment

Fuzzy RD Is IV

Think of the running variable as an **instrument**

- We introduce a new dummy variable T_i, which indicates whether the running variable has crossed the cutoff:

$$T_i = \begin{cases} 1 & \text{if} \quad X_i \geq c \\ 0 & \text{if} \quad X_i < c \end{cases}$$

- With fuzzy RD, we have $D_i \neq T_i$ for some individuals

- T_i is an instrument for D_i in a regression model for Y_i on X_i and D_i

Fuzzy RD Is IV

We can define an estimator that is in the spirit of IV:

$$\tau_{FRD} = \frac{\rho}{\phi} = \frac{\text{Effect of threshold on } Y_i}{\text{Effect of threshold on } D_i}$$

- Jump at the cutoff in the outcome needs to be rescaled by the jump at the cutoff in the probability of treatment
- Fuzzy RD estimates the local average treatment effect (LATE) at the threshold for compliers

Fuzzy RD Estimation

First stage:

$$D_i = \alpha_1 + \beta_1 X_i + \phi T_i + \epsilon_{1i}$$

Second stage (notice the fitted values $\widehat{D}_i$):

$$Y_i = \alpha_2 + \beta_2 X_i + \tau_{FRD} \widehat{D}_i + \epsilon_{2i}$$

Reduced form:

$$Y_i = \alpha_0 + \beta_0 X_i + \rho T_i + \epsilon_{0i}$$

$\rho = \phi \tau_{FRD}$ can be interpreted as an intention-to-treat (ITT) effect

Peer Effects in School

Many parents who are looking for a home are willing to pay a premium to have their children in “better” schools

- Parents and teachers believe that peers matter: Having higher-achieving classmates improves own learning
- Regression that controls for own past achievement suggests strong peer effects (~0.25)
- But students from the same class/school tend to be similar in many ways (e.g. family background) → selection bias

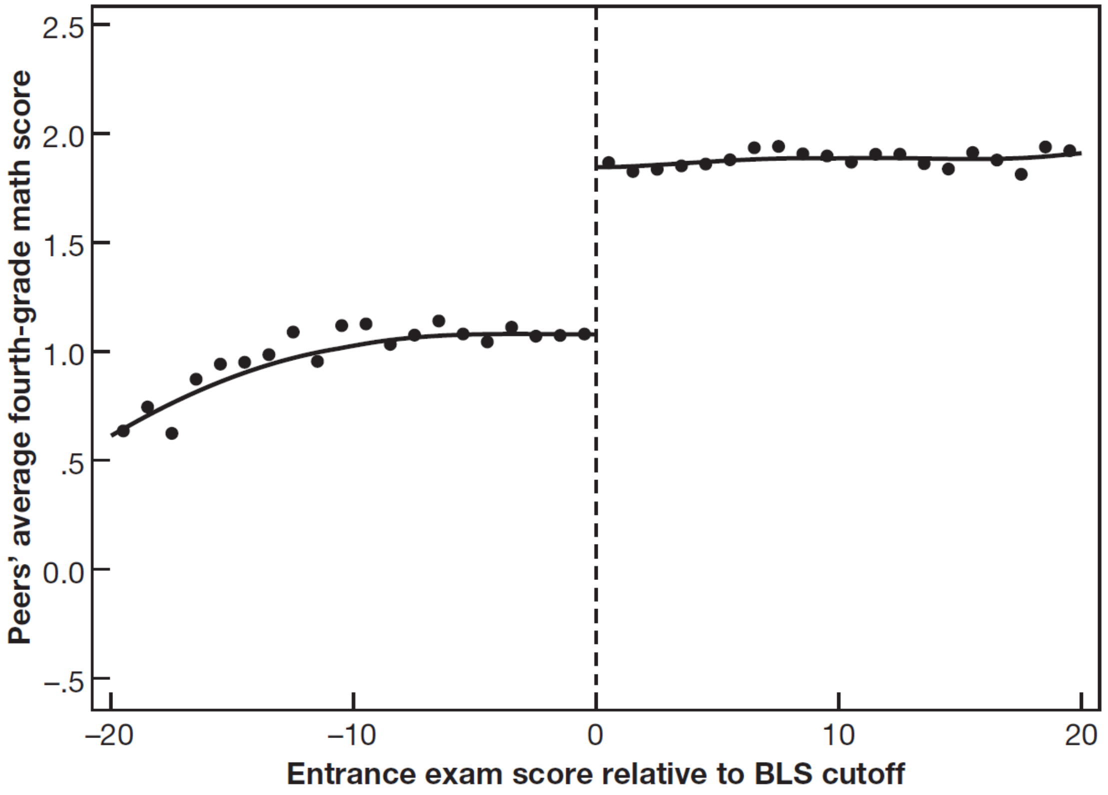
Peer Effects in School

Exam schools offer public school students the opportunity to attend schools with higher achieving peers

- Students are selected by an admissions test with sharp cutoffs
- Applicants who qualify for a top exam school attend schools with higher-achieving peers → Fuzzy RD
- Here we consider a fuzzy RD design where the intensity rather than probability of treatment (i.e. peer quality) jumps

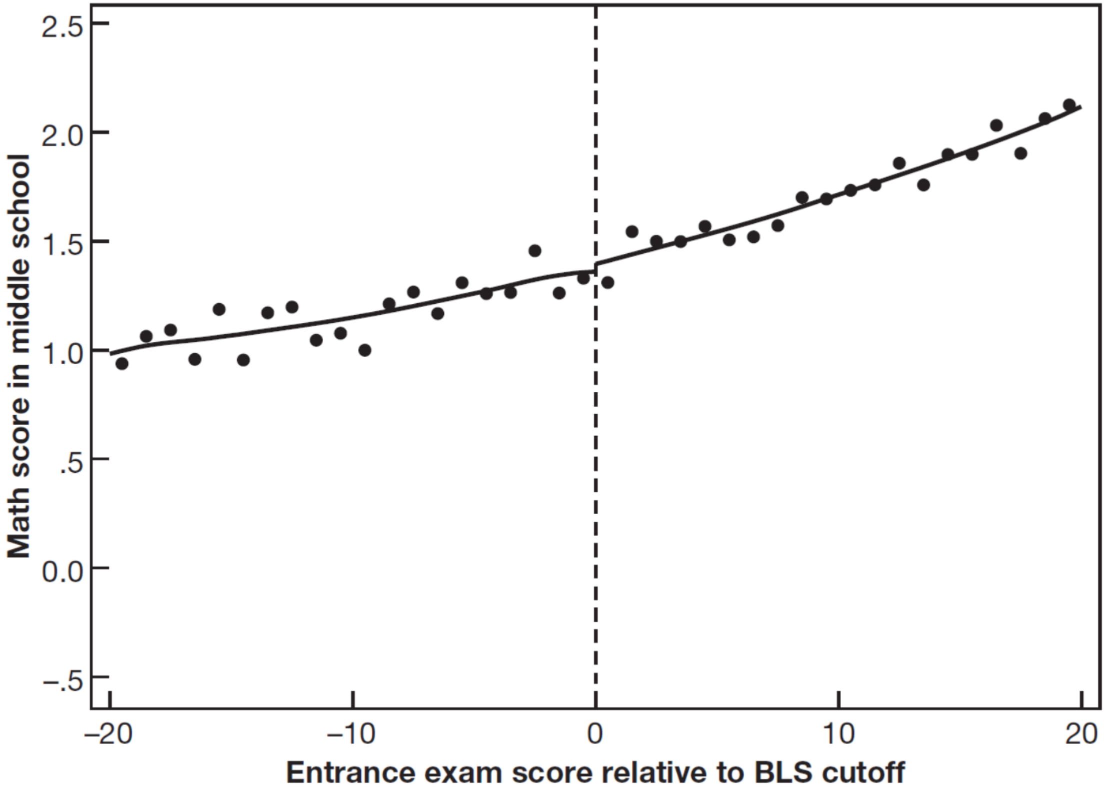
First Stage

Peer quality is 0.8 SD higher for applicants who are just above the admissions cutoff of a top exam school (e.g. Boston Latin School):



Reduced Form

No similar jump in own achievement:



Reduced Form Regression

$$Y_i = \alpha_0 + \beta_0 X_i + \rho T_i + \epsilon_{0i}$$

- Y_i is student i's 7th grade math score
- T_i is a dummy that equals to 1 if student i qualifies for a top exam school (and 0 otherwise)
- X_i is the running variable for student i's score in the entry exam

$\hat{\rho} = -0.02$ ($SE_{\hat{\rho}} = 0.10$) → jump is not significantly different from zero

2SLS Regression

First stage: $\overline{D}_{(i)} = \alpha_1 + \beta_1 X_i + \phi T_i + \epsilon_{1i}$

- $\overline{D}_{(i)}$ is the average 4th grade math score of student i's classmates (i.e. average peer quality)

Second stage: $Y_i = \alpha_2 + \beta_2 X_i + \tau_{FRD} \widehat{\overline{D}}_i + \epsilon_{2i}$

- Y_i is the 7th grade math score of student i

$\hat{\tau}_{FRD} = -0.023$ ($SE_{\hat{\tau}_{FRD}} = 0.132$) → no significant peer effects

Threats to Identification in RD Designs

RD designs assume individuals have imperfect control over the running variable

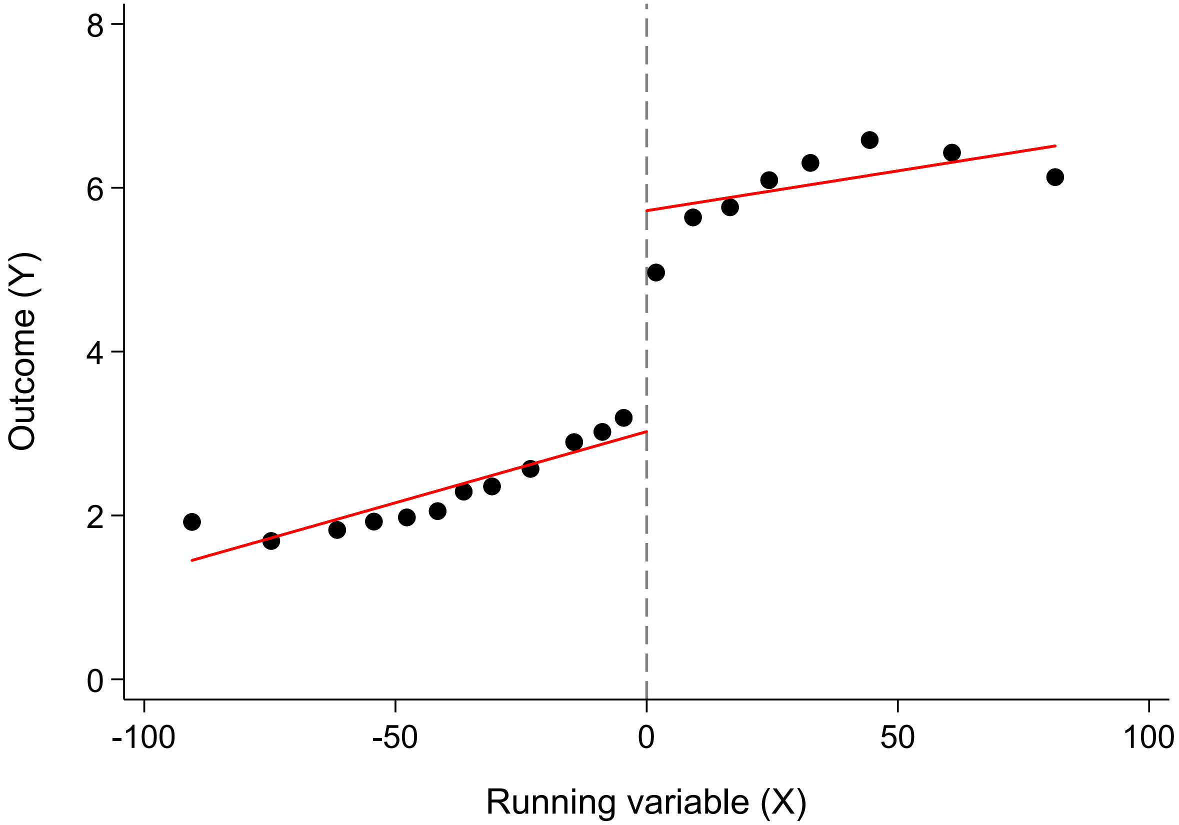
Individuals may be "gaming" the threshold:

- Check whether there is bunching of individuals on either side of the cutoff (McCrary density test)
- While not perfect, you can also test for covariate balance around the cutoff
- Perform placebo tests at arbitrary cutoffs (where there shouldn't be any effects)

Lee and Lemieux (2010)

RD Visualization

Present the main graphs using binned local averages:



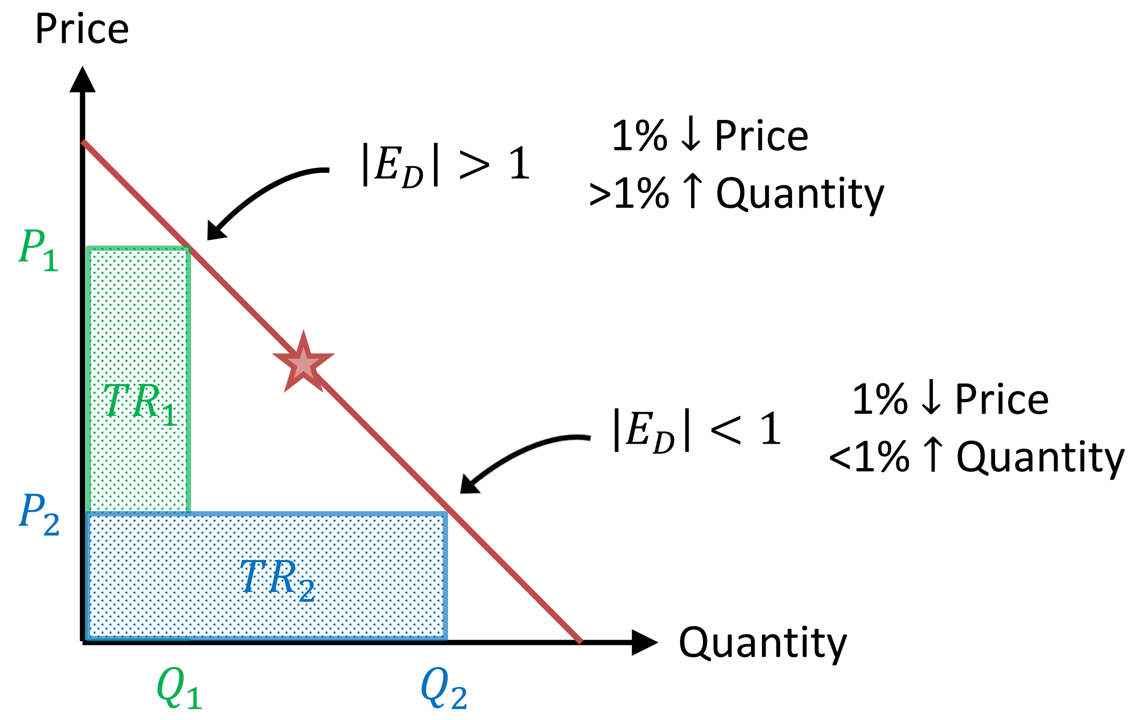
SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Credits:
Alain Cohn
Assistant Professor of Information

Differences-in-Differences

Example – Pricing Strategy

- Imagine you want to know whether you should raise or lower the price to increase revenue



Emily Glassberg Sands: https://medium.com/teconomics-blog/5-tricks-when-ab-testing-is-off-the-table-f2637e9f15a5

Example – Pricing Strategy

- How can we learn where we are on the demand curve?
- Randomizing the price for each user may not be feasible
- Change price in some regions (e.g. countries), but not others
  - Measure sales for treated and control groups *before* and *after* the price change
  - When units are observed across multiple points in time → **panel or longitudinal data**

Basic DD Setup

- Two groups, two time periods (t = 0, t = 1)
- Neither group is treated at $t = 0$
- Treatment group is treated at $t = 1$, but the control group is not
- Change in the treatment group from $t = 0$ to $t = 1$ might be correlated with a time trend in the outcome
- Change in the control group from $t = 0$ to $t = 1$ identifies the time trend

DD Estimator

- **Differences-in-differences (DD)** compares the change before and after the intervention in the treatment group with the same change in the control group:

$$\tau_{DD} = (Y_{T1} - Y_{T0}) - (Y_{C1} - Y_{C0})$$

- Without randomization, the difference-in-differences is an estimate of the **average treatment effect on the treated (ATT)**

2 Periods

Month	Group		
	Control (Canada)	Treat (U.S.)	Total
May	76.9	91.5	168.4
Jun	69.3	107.9	177.3
Total	146.3	199.4	345.7

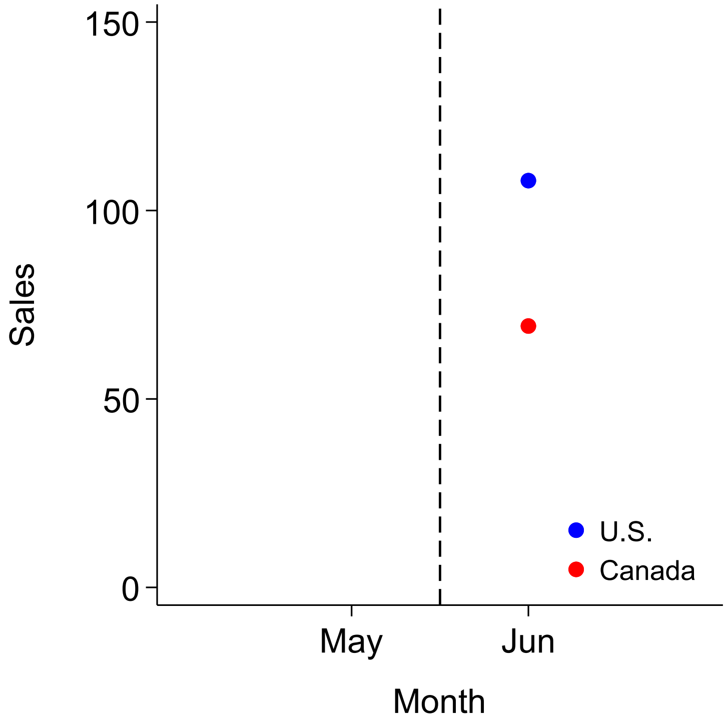
In June U.S. price is lowered; no reduction in Canada

How can we learn about the causal effect of a price reduction on revenue?

Simple Comparison Between Groups

Month	Group Control	Treat	Total
May	76.9	91.5	168.4
Jun	69.3	107.9	177.3
Total	146.3	199.4	345.7

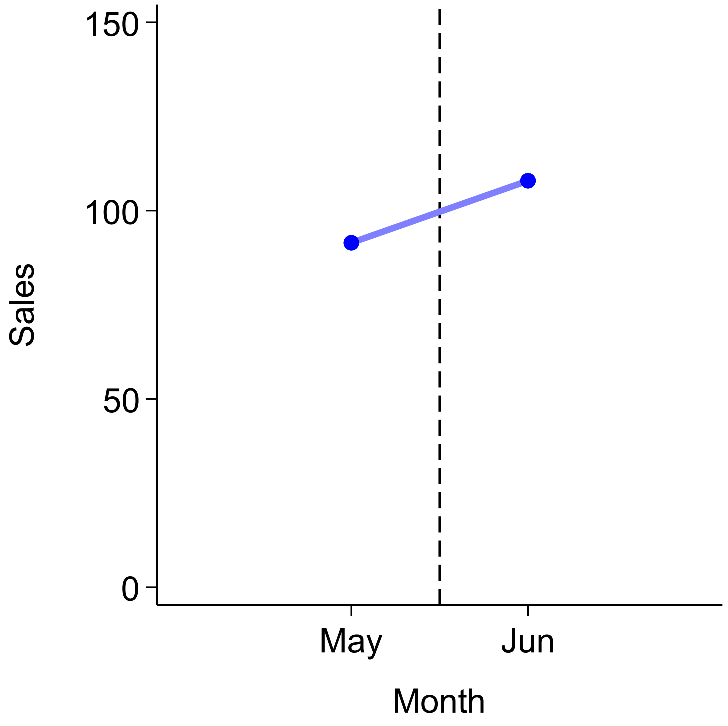
$$107.9 - 69.3 = 38.6$$



Before-After Comparison for Treatment Group

Month	Group: Control	Treat	Total
May	76.9	91.5	168.4
Jun	69.3	107.9	177.3
Total	146.3	199.4	345.7

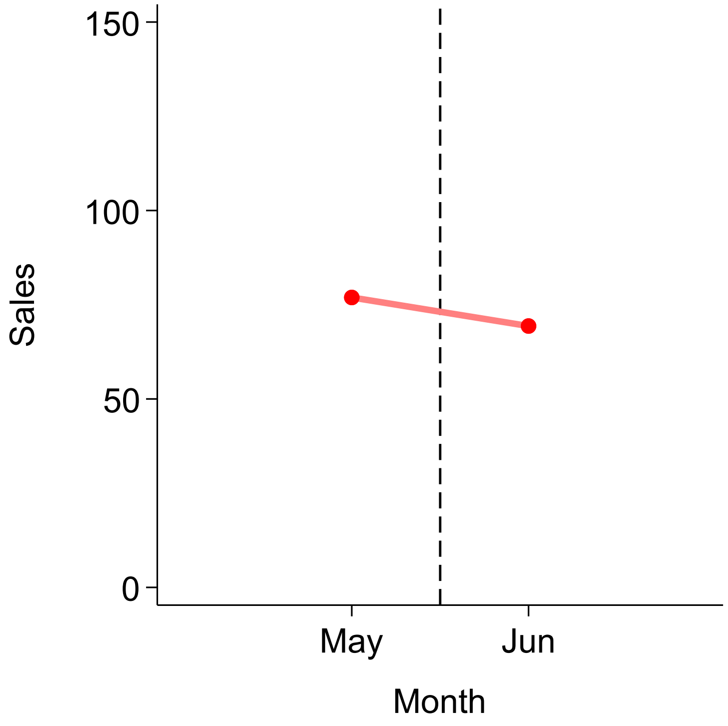
$$107.9 - 91.5 = 16.4$$



Before-After Comparison for Control Group

Month	Control	Group Treat	Total
May	76.9	91.5	168.4
Jun	69.3	107.9	177.3
Total	146.3	199.4	345.7

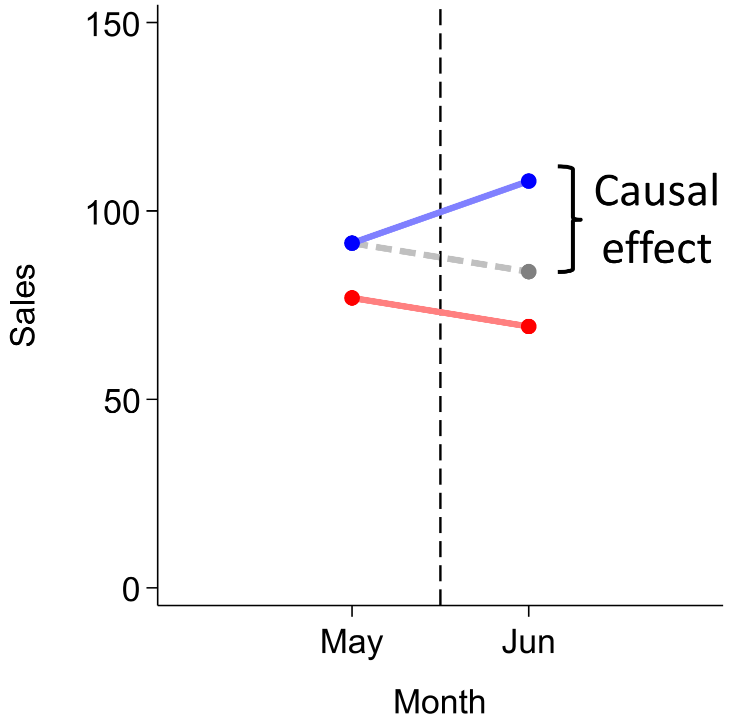
$$76.9 - 69.3 = -7.6$$



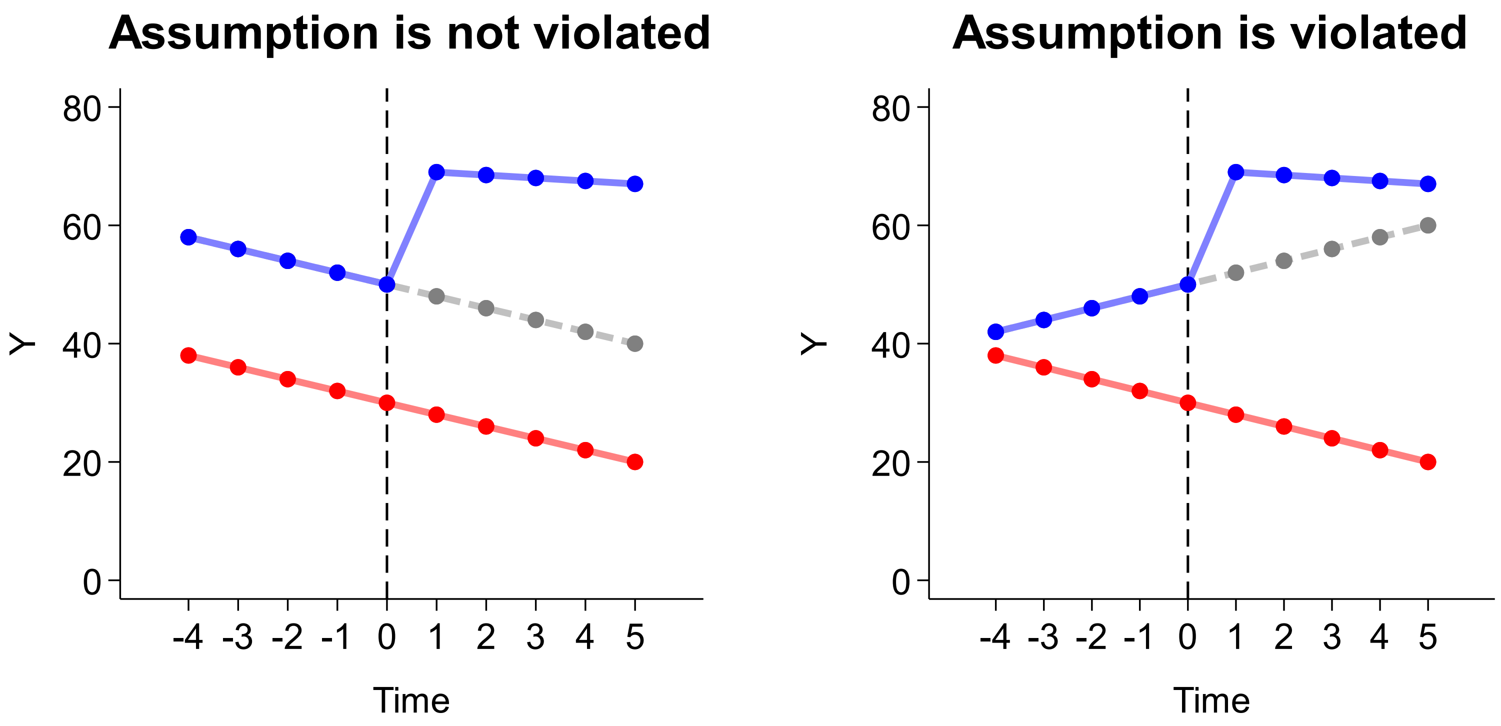
Difference-in-Differences

Month	Group Control	Treat	Total
May	76.9	91.5	168.4
Jun	69.3	107.9	177.3
Total	146.3	199.4	345.7

$$(107.9 - 91.5) - (69.3 - 76.9)$$
$$= (16.4) - (-7.6) = 24.0$$



Parallel Trends Assumption



- Parallel trend

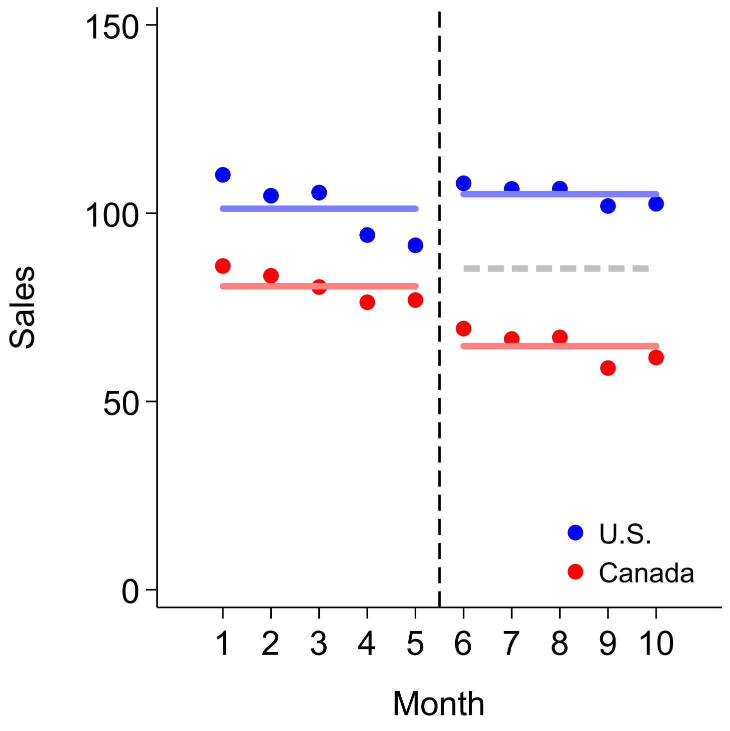
SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Credits:
Alain Cohn
Assistant Professor of Information

Differences-in-Differences Regression

More than 2 Periods

- Revenue from Jan to Oct (price reduction from Jun to Oct in the U.S.):



DD Regression

- Typical DD regression model:

$$Y_{it} = \alpha + \beta\, TREAT_i + \gamma\, POST_t + \tau_{DD}\, (TREAT \times POST)_{it} + \epsilon_{it}$$

  - Y_{it} is the outcome for unit i at time t
  - $TREAT_i$ is a dummy equal to 1 if units are treated in the post-treatment period
  - $POST_t$ is a post-treatment dummy
  - $(TREAT \times POST)_{it}$ is an interaction term equal to 1 for the treatment group in the post-treatment period

DD Regression

- Typical DD regression model:

$$Y_{it} = \alpha + \beta\, TREAT_i + \gamma\, POST_t + \tau_{DD}\, (TREAT \times POST)_{it} + \epsilon_{it}$$

  - CAN Pre: α
  - CAN Post: $\alpha + \gamma$
  - US Pre: $\alpha + \beta$
  - US Post: $\alpha + \beta + \gamma + \tau_{DD}$

- DD estimate: (US Post − US Pre) − (CAN Post − CAN Pre) $= \tau_{DD}$

DD Regression

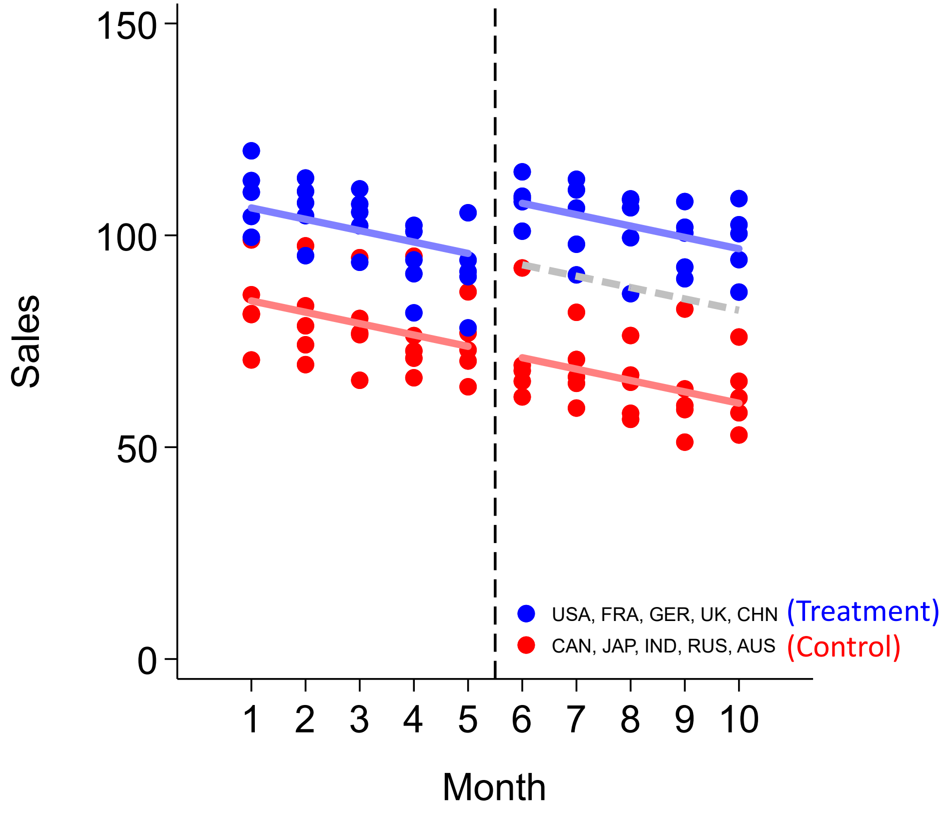
```
. regress y treat post treatXpost, robust

Linear regression                               Number of obs     =         20
                                                F(3, 16)          =     121.88
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9259
                                                Root MSE          =     5.1542
```

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
treat	20.59374	4.017559	5.13	0.000	12.07689	29.11058
post	-15.89622	2.66304	-5.97	0.000	-21.54161	-10.25083
treatXpost	19.76166	4.610078	4.29	0.001	9.988728	29.53459
_cons	80.59598	1.847217	43.63	0.000	76.68006	84.51191

- `treat`: U.S. site has generally higher revenue (+20.6)
- `post`: revenue is lower in the post-treatment period (-15.9)
- `treatXpost`: price reduction increases revenue (+19.8)

Multiple Groups, Multiple Periods

- 5 treatment and 5 control groups (priced reduced in treat markets from Jun to Oct):



DD Regression with Fixed Effects

$$Y_{it} = \alpha + \sum_{k=Country\ \#2}^{Country\ \#10} \beta_k\, COUNTRY_{ki} + \sum_{j=Feb}^{Oct} \boxed{\gamma_j} MONTH_{jt} + \tau_{DD}(TREAT \times POST)_{it} + \epsilon_{it}$$

- $COUNTRY_{ki}$ is a set of dummies for each country
- β_k capture time-invariant country fixed effects
- $MONTH_{jt}$ is a set of dummies for each month
- γ_j capture time effects that are common to all countries

• One group and one time fixed effect serves as the reference group (captured by α) and has to be omitted

DD Regression with Fixed Effects

```
. reghdfe y treatXpost, vce(cluster country) absorb(i.id i.t )
(MWFE estimator converged in 2 iterations)

HDFE Linear regression                          Number of obs   =        100
Absorbing 2 HDFE groups                         F(   1,      9) =      97.90
Statistics robust to heteroskedasticity         Prob > F        =     0.0000
                                                R-squared       =     0.9772
                                                Adj R-squared   =     0.9715
                                                Within R-sq.    =     0.6156
Number of clusters (country) =         10       Root MSE        =     3.0369

                              (Std. Err. adjusted for 10 clusters in country)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  treatXpost |    13.6642   1.381029     9.89   0.000      10.5401    16.78831
.      _cons |   83.63633   .3452573   242.24   0.000      82.8553    84.41735
------------------------------------------------------------------------------
```

– Output of country and month fixed effects is suppressed

Standard Errors in DD

- Problem of serial correlation in DD applications:
  - Many DD applications use data from many periods (not just 1 pre and 1 post period)
  - Conventional standard errors tend to be downward biased because of serial correlation in the outcome variable (Bertrand et al. 2004)
- Solutions:
  - Clustered standard errors at the group level (>30 groups)
  - Cluster (or block) bootstrap standard errors (Cameron et al. 2008)
  - Collapse time series into 1 pre and 1 post period for each group

Threats to Identification in DD Designs

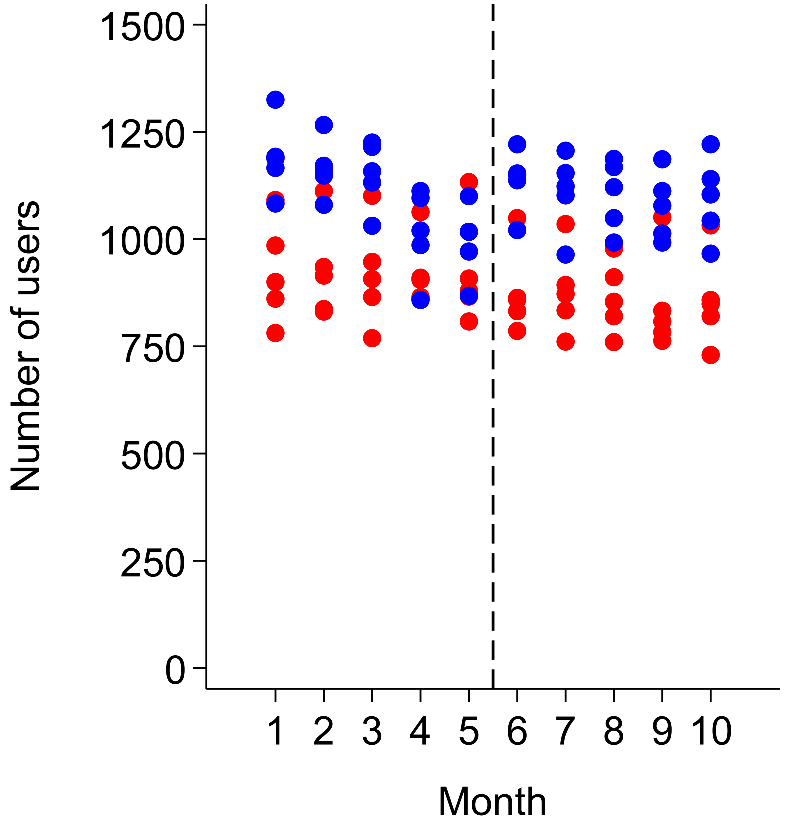
- Intervention is *not* as good as random
- Targeting based on pre-existing differences in outcomes
- E.g. price change in most promising markets

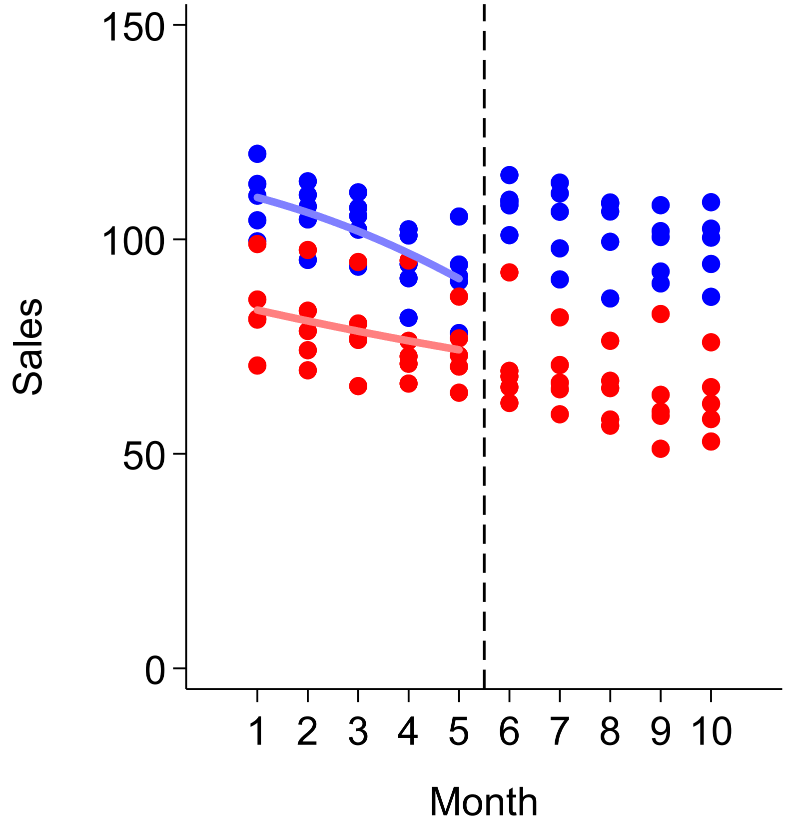
- Some units are more likely to select into the treatment
- "Ashenfelter dip": people who participate in job training programs often experience a dip in earnings just before they enter the program (which is why they enroll)

- Targeting or selection → parallel trends assumption might be violated

Non-Parallel Trends?

- Dip in number of users in treat markets just before the price reduction is implemented:





Regression DD with Covariates

- Parallel trends assumption may hold conditional on covariates

- Add covariates in a linear, additive way:

$$Y_{it} = \alpha + \sum^{k} \beta_k \, COUNTRY_{ki} + \sum^{j} \gamma_j \, MONTH_{jt} + \delta \boldsymbol{X_{it}} + \tau_{DD}(TREAT \times POST)_{it} + \epsilon_{it}$$

  – X_{it} is the number of users in country i and month t

Regression DD with Covariates

```
. reghdfe y treatXpost user, vce(cluster country) absorb(i.id i.t )
(MWFE estimator converged in 2 iterations)

HDFE Linear regression                            Number of obs     =        100
Absorbing 2 HDFE groups                           F(   2,        9) =     112.46
Statistics robust to heteroskedasticity           Prob > F          =     0.0000
                                                  R-squared         =     0.9869
                                                  Adj R-squared     =     0.9834
                                                  Within R-sq.      =     0.7794
Number of clusters (country) =         10         Root MSE          =     2.3156

                              (Std. Err. adjusted for 10 clusters in country)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  treatXpost |   10.94872   1.059607    10.33   0.000     8.551718    13.34571
        user |   .0478416   .0039587    12.09   0.000     .0388864    .0567968
        cons |   36.54201   3.995078     9.15   0.000     27.50452    45.57951
------------------------------------------------------------------------------
```

- `user`: revenue increases by 0.05 for each additional user
- `treatXpost`: estimate is robust to controlling for the number of users

Unit-Specific Linear Time Trends

- With more than two periods, we can add unit-specific *linear* time trends to the regression model:
- $$Y_{it} = \alpha + \sum^{k} \beta_k \, COUNTRY_{ki} + \sum^{j} \gamma_j \, MONTH_{jt} + \delta X_{it}$$
  $$+ \tau_{DD}(TREAT \times POST)_{it} + \sum^{k} \theta_k \boxed{(\boldsymbol{COUNTRY_{ki}} \times \boldsymbol{t})} + \epsilon_{it}$$
  – $t = 1, 2, \dots, T$ is a linear time trend that increases by one for each month
- While this relaxes the parallel trends assumption, we assume that the trends continue post treatment

Unit-Specific Linear Time Trends

```
. reghdfe y  treatXpost user tXcountry2-tXcountry10, vce(cluster country) absorb(i.id i.t )
(MWFE estimator converged in 2 iterations)

HDFE Linear regression                            Number of obs     =        100
Absorbing 2 HDFE groups                           F(  11,        9) =          .
Statistics robust to heteroskedasticity           Prob > F          =          .
                                                  R-squared         =     0.9886
                                                  Adj R-squared     =     0.9837
                                                  Within R-sq.      =     0.8078
Number of clusters (country) =         10         Root MSE          =     2.2978

                              (Std. Err. adjusted for 10 clusters in country)
```

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
treatXpost	11.24196	1.886112	5.96	0.000	6.97528	15.50865
user	.0452953	.0026831	16.88	0.000	.0392258	.0513649
tXcountry2	.5796031	.0171879	33.72	0.000	.5407213	.6184849
tXcountry3	.0161571	.0204401	0.79	0.450	-.0300817	.0623959
tXcountry4	.1772667	.0100005	17.73	0.000	.1546439	.1998895
tXcountry5	.8173548	.0160334	50.98	0.000	.7810847	.8536248
tXcountry6	.2218789	.2857755	0.78	0.457	-.4245901	.8683479
tXcountry7	.3465776	.2862734	1.21	0.257	-.3010179	.9941732
tXcountry8	.129999	.2864311	0.45	0.661	-.5179531	.7779512
tXcountry9	.5799866	.2874191	2.02	0.074	-.0702006	1.230174
tXcountry10	.163218	.2861734	0.57	0.582	-.4841512	.8105873
_cons	37.3437	2.616624	14.27	0.000	31.42449	43.26291

– `tXcountry*`: country-specific linear time trends

Robustness Checks

- What can we do to check for robustness of DD results?
  - Include leads of the treatment (as if the treatment started in a prior period)
  - Use alternative control groups that experience the same unit-specific time trends
  - Use alternative outcomes that are not supposed to be affected by the treatment

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Credits:
Alain Cohn
Assistant Professor of Information