Microeconometrics

Counterfactuals: from ATE to LATE

Alex Armand
alex.armand@novasbe.pt

Counterfactual

Alternative scenarios we can't directly observe but must infer.

The government introduced the new job training program six month ago, and today unemployment is down by 10%.

A city banned plastic bags and the litter reduced by 25% as compared to other cities.

A study finds that people who eat chocolate live longer than those who do not.

Understanding counterfactuals ensures that causal claims are well-founded.

Counterfactual

Consider the case in which we **compare groups** of individuals

- Assume **2 groups**:
  1. **Group D** eats chocolate.
  2. **Group C** does not eat chocolate.

- This is what we are doing in **linear models** to compare groups:

$$y_i = \alpha + \beta d_i + \epsilon_i$$

$$d_i = \begin{cases} 1 & i \text{ is in group D} \\ 0 & i \text{ is in group C} \end{cases}$$

- What are α and β?
- What assumptions we need to identify β?

The potential outcomes model

Neyman - Fisher - Cox - Roy - Quandt - Rubin model

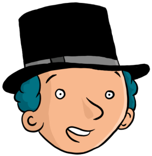
- y_i is an **outcome of interest** for individual i
- d_i is a **group indicator** (1 if the individual i is in group D and 0 if in group C)
- **Potential outcomes** for individual i: denoted by

$$\begin{aligned} y_{1i} &= \beta + \alpha_i + u_{1i} \quad \text{if } d_i = 1 \\ y_{0i} &= \beta + u_{0i} \qquad\qquad \text{if } d_i = 0 \end{aligned}$$

  - α_i is the effect of the treatment
  - **Stable Unit Treatment VAlue (SUTVA) assumption**: y_{1i}, y_{0i} and d_i don't depend on j

The potential outcomes model: a simple example

How John react when offered **additional pocket money**?



additional pocket money  (P)

on John's consumption of sweets  (Y)?

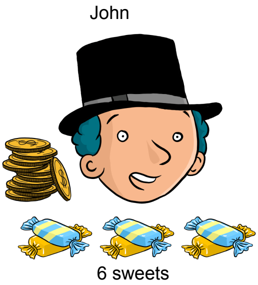
Can we observe y_{1i} and y_{0i} for John?

The treatment model: potential versus observed

From the data, **we observe y_i**

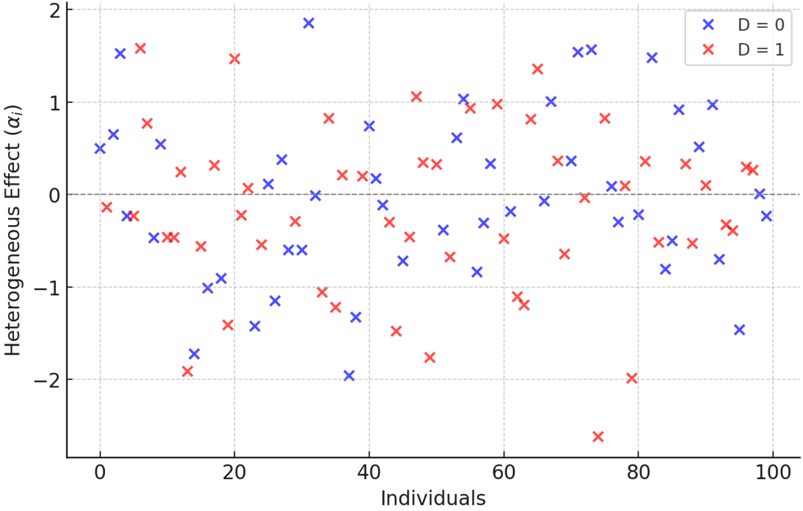
$$
\begin{aligned}
y_i &= y_{1i}d_i + y_{0i}(1 - d_i) = y_{0i} + (y_{1i} - y_{0i})d_i \\
&= y_{0i} + \alpha_i d_i
\end{aligned}
$$

- Effects are **heterogeneous**: α_i is individual-specific!
- **Fundamental observability problem**: we only observe one of the two potential outcomes



What can we learn about α_i?

Ideally, we would like to observe α_i for each individual.



What can we learn about α_i?

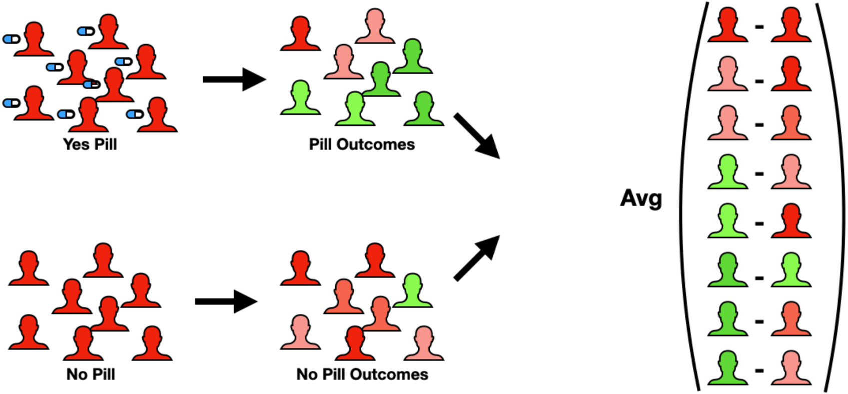
- Estimation methods typically do not identify α_i

$$y_i = \beta + \alpha_i d_i + u_i$$

- **Only some average of this parameter** over some (sub-)population:
  - $E[\alpha_i]$: average treatment effect (**ATE**)
  - $E[\alpha_i | d_i = 1]$: average effect on individuals that were assigned to treatment (**ATT**)
  - $E[\alpha_i | d_i = 0]$: a average effect on non-participants (**ATNT**)
  - $E[\alpha_i | z = z*]$: a local average of the effect (**LATE**)

Causal inference comparing groups

An example from medical studies:



Causal inference comparing groups

- In the sample, the following averages of y_i can be computed:
  - Average for **Yes Pill**: $E[y_i|d_i = 1]$
  - Average for **No Pill**: $E[y_i|d_i = 0]$
- **Take the difference**:

$$
\begin{aligned}
E[y_i|d_i = 1] - E[y_i|d_i = 0] &= E[y_{1i}|d_i = 1] - E[y_{0i}|d_i = 0] \\
&= E[y_{1i} - y_{0i}|d_i = 1] + \\
&\quad \{E[y_{0i}|d_i = 1] - E[y_{0i}|d_i = 0]\}
\end{aligned}
$$

- Difference in means is equal to **ATT + {selection bias}!**
- Intuition:
  - ATT: effect of pill in the Yes Pill group
  - Selection bias: difference in outcomes driven by characteristics of Yes Pill individuals

Causal inference comparing groups

- In the sample, the following averages of y_i can be computed:
  - Average for **Yes Pill**: $E[y_i|d_i = 1]$
  - Average for **No Pill**: $E[y_i|d_i = 0]$
- **Take the difference**:

$$
\begin{aligned}
E[y_i|d_i = 1] - E[y_i|d_i = 0] &= E[y_{1i}|d_i = 1] - E[y_{0i}|d_i = 0] \\
&= E[y_{1i} - y_{0i}|d_i = 1] + \\
&\quad \{E[y_{0i}|d_i = 1] - E[y_{0i}|d_i = 0]\}
\end{aligned}
$$

- Difference in means is equal to **ATT + {selection bias}!**
- Intuition:
  - ATT: effect of pill in the Yes Pill group
  - Selection bias: difference in outcomes driven by characteristics of Yes Pill individuals

Back into OLS

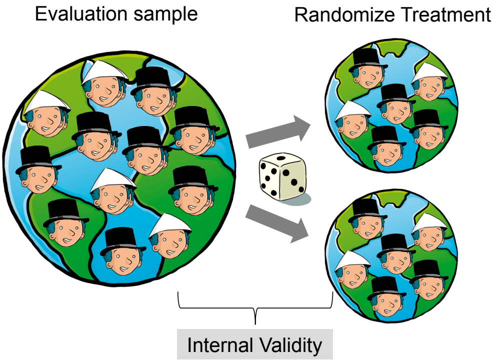
When we estimate the following model:

$$y_i = \beta + \alpha d_i + \epsilon_i$$

- **How can we estimate this model?**
  - Orthogonality fails when:
    1. **Selection on the observables**: ϵ_i contains observable characteristics that determine d_i
    2. **Selection on the unobservables**: ϵ_i contains unobservable characteristics that determine d_i

The social experiment

Assignment to treatment is at random: groups are equal in all aspects apart from the treatment status.



- Random assignment determines the following assumptions:
  1. **R1**: $E[u_i|d_i = 1] = E[u_i|d_i = 0] = E[u_i]$
  2. **R2**: $E[\alpha_i|d_i = 1] = E[\alpha_i|d_i = 0] = E[\alpha_i]$

Comparing means in the social experiment

- Under R1 and R2: $E[y_{1i}|d_i = 1] = E[y_{1i}|d_i = 0] = E[y_{1i}]$
- Comparing means we obtain (recall from ATT + bias)

$$\begin{aligned} E[y_i|d_i = 1] - E[y_i|d_i = 0] &= E[y_{1i} - y_{0i}|d_i = 1] + \\ &\quad \{E[y_{0i}|d_i = 1] - E[y_{0i}|d_i = 0]\} \\ &= E[y_{1i} - y_{0i}|d_i = 1] \\ &= E[y_{1i} - y_{0i}] \end{aligned}$$

- The difference is identified with an **OLS regression of the treatment indicator on the outcome variable** using the cross-section post-treatment!

$$y_i = \beta + \alpha_{ATE} d_i + u_i$$

APPLICATION: Lalonde (1986) dataset

We will make use of the following paper: Lalonde, R.J. (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", American Economic Review, 76, 604-620.

```
  obs:         3,509                          NSW: treated and control groups
 vars:            11                          24 Oct 2012 10:31
```

variable name	storage type	display format	value label	variable label
treated	byte	%16.0g	treated	NSW treated (1), NSW controls (0)
age	byte	%9.0g		Age
age2	int	%9.0g		Age (squared)
educ	byte	%9.0g		Schooling (years)
black	byte	%9.0g	dummy	Black
hispanic	byte	%9.0g	dummy	Hispanic
married	byte	%9.0g	dummy	Married
nodegree	byte	%9.0g	dummy	<12 years of education
re75	float	%9.0g		Real earnings (1975)
re78	float	%9.0g		Real earnings (1978)
randomized	float	%9.0g	sample	NSW sample (1) PSID sample (0)

Experimental vs non-experimental data

Combines **cross-sections** data from two different populations:

1. **experimental**: National Supported Work (NSW) programme
   - Employment program designed to help disadvantaged workers
   - NSW was **assigning applicants to available positions at random**

2. **non-experimental**: Panel Study of Income Dynamics (PSID) dataset
   - Sample representative of the working-age population

```
.         tabulate treated
```

NSW treated (1), NSW controls (0)	Freq.	Percent	Cum.
-	2,915	83.07	83.07
Treated	594	16.93	100.00
Total	3,509	100.00	

Experimental dataset

- Drop the observations for which the variable *randomization* equals 0.
- The first step in a social experiment is to check balance across control and treatment group
  1. t-test on each of the variables
  2. Hotelling T-squared test of the hypothesis that the vector of means of all variables are equal across groups
- If randomization is confirmed, then we can apply OLS for estimating ATE using the cross-section

Comparing balance for individual variables

Performing a **t-test on individual variables** allows

- Comparing equality at the mean across the two groups
- Identify variables to use as control variable in OLS

```
.                         ttest age,   by(treat) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
-	425	24.44706	.3196754	6.590276	23.81871	25.0754
Treated	297	24.62626	.3879837	6.686391	23.86271	25.38982
combined	722	24.52078	.2465922	6.625947	24.03665	25.0049
diff		-.1792038	.5027163		-1.166403	.807995

```
    diff = mean(-) - mean(Treated)                                t =  -0.3565
Ho: diff = 0                     Satterthwaite's degrees of freedom =  631.223

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3608         Pr(|T| > |t|) = 0.7216          Pr(T > t) = 0.6392
```

Comparing balance for individual variables

Performing a **t-test on individual variables** allows

- Comparing equality at the mean across the two groups
- Identify variables to use as control variable in OLS

```
.                          ttest educ,  by(treat) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
-	425	10.18824	.0785178	1.618686	10.0339	10.34257
Treated	297	10.38047	.1054743	1.817712	10.1729	10.58805
combined	722	10.26731	.0634451	1.704774	10.14275	10.39187
diff		-.1922361	.131491		-.4504846	.0660124

```
    diff = mean(-) - mean(Treated)                                t =  -1.4620
Ho: diff = 0                         Satterthwaite's degrees of freedom =  588.748

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0721         Pr(|T| > |t|) = 0.1443          Pr(T > t) = 0.9279
```

Comparing balance for individual variables

Performing a **t-test on individual variables** allows

- Comparing equality at the mean across the two groups
- Identify variables to use as control variable in OLS

```
.                          ttest black, by(treat) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       - |     425          .8    .0194257    .4004714    .7618173    .8381827
 Treated |     297    .8013468    .0231906    .3996597    .7557074    .8469862
---------+--------------------------------------------------------------------
combined |     722     .800554    .0148813    .3998609    .7713382    .8297698
---------+--------------------------------------------------------------------
    diff |           -.0013468    .0302517               -.0607517    .0580581
------------------------------------------------------------------------------
    diff = mean(-) - mean(Treated)                                t =  -0.0445
Ho: diff = 0                     Satterthwaite's degrees of freedom =  637.876

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.4823         Pr(|T| > |t|) = 0.9645          Pr(T > t) = 0.5177
```

Comparing balance for individual variables

Performing a **t-test on individual variables** allows

- Comparing equality at the mean across the two groups
- Identify variables to use as control variable in OLS

```
.                          ttest re75,  by(treat) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       - |     425    3026.683    252.2977     5201.25    2530.773    3522.593
 Treated |     297    3066.098    282.8697    4874.889    2509.407    3622.789
---------+--------------------------------------------------------------------
combined |     722    3042.897    188.5423    5066.143    2672.739    3413.054
---------+--------------------------------------------------------------------
    diff |           -39.41544    379.0375               -783.6763    704.8454
------------------------------------------------------------------------------
    diff = mean(-) - mean(Treated)                                t =  -0.1040
Ho: diff = 0                     Satterthwaite's degrees of freedom =  661.861

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.4586         Pr(|T| > |t|) = 0.9172          Pr(T > t) = 0.5414
```

Comparing overall balance with Hotelling test

1. Run OLS of all variables on treatment indicator

```
. reg treat age educ black hispanic married nodegree re75

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(7, 714)       =      1.13
       Model |  1.91497145         7  .273567349   Prob > F        =    0.3423
    Residual |  172.911898       714  .242173527   R-squared       =    0.0110
-------------+----------------------------------   Adj R-squared   =    0.0013
       Total |   174.82687       721  .242478322   Root MSE        =    .49211

------------------------------------------------------------------------------
     treated |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0003867   .0028944    -0.13   0.894    -.0060693    .0052959
        educ |  -.0056696   .0143623    -0.39   0.693    -.0338669    .0225277
       black |   -.023763   .0640981    -0.37   0.711    -.1496064    .1020803
    hispanic |  -.0602687   .0836427    -0.72   0.471    -.2244838    .1039464
     married |    .022314    .052165     0.43   0.669    -.0801011    .1247291
    nodegree |  -.1295037   .0592253    -2.19   0.029    -.2457803    -.013227
        re75 |  -7.54e-07   3.73e-06    -0.20   0.840    -8.09e-06    6.58e-06
       _cons |   .6040808   .2070859     2.92   0.004     .1975107    1.010651
------------------------------------------------------------------------------
```

2. Test joint significance of all variables (constant excluded)

```
. test  age educ black hispanic married nodegree re75

 ( 1)  age = 0
 ( 2)  educ = 0
 ( 3)  black = 0
 ( 4)  hispanic = 0
 ( 5)  married = 0
 ( 6)  nodegree = 0
 ( 7)  re75 = 0

       F(  7,   714) =    1.13
            Prob > F =    0.3423
```

Estimate impact with OLS: no controls

Positive effect (significant at 10%) - notice for simplicity we assume homosckedasticity

```
.                          regress re78 treated

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(1, 720)       =      3.52
       Model |   137332501         1   137332501   Prob > F        =    0.0609
    Residual |  2.8053e+10       720  38962866.3   R-squared       =    0.0049
-------------+----------------------------------   Adj R-squared   =    0.0035
       Total |  2.8191e+10       721  39099301.3   Root MSE        =      6242

------------------------------------------------------------------------------
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |   886.3037   472.0863     1.88   0.061    -40.52635    1813.134
       _cons |   5090.048   302.7826    16.81   0.000     4495.606    5684.491
------------------------------------------------------------------------------
```

Estimate impact with OLS: controls

Introducing controls reduces slightly the size of the effect (still significant at 10%) – **why?**

```
. regress re78 treated age age2 educ black hispanic nodegree
```

Source	SS	df	MS
Model	670296792	7	95756684.6
Residual	2.7520e+10	714	38543836.8
Total	2.8191e+10	721	39099301.3

Number of obs	=	722
F(7, 714)	=	2.48
Prob > F	=	0.0159
R-squared	=	0.0238
Adj R-squared	=	0.0142
Root MSE	=	6208.4

re78	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
treated	798.3512	472.1283	1.69	0.091	-128.5747	1725.277
age	-3.805475	211.1663	-0.02	0.986	-418.3866	410.7756
age2	.5296508	3.556177	0.15	0.882	-6.452164	7.511466
educ	219.7946	182.9296	1.20	0.230	-139.3496	578.9387
black	-1762.833	803.88	-2.19	0.029	-3341.084	-184.5814
hispanic	-117.148	1054.228	-0.11	0.912	-2186.906	1952.61
nodegree	-494.2816	749.2561	-0.66	0.510	-1965.29	976.727
_cons	4430.163	3653.224	1.21	0.226	-2742.183	11602.51

Estimate impact with OLS: heterogeneity

Example: estimate impact for younger (less than and older than 24 y.o.)

```
.                    regress re78 treated if age <= 24

      Source |       SS           df       MS      Number of obs   =       408
-------------+----------------------------------   F(1, 406)       =      0.39
       Model |  11632102.9         1  11632102.9   Prob > F        =    0.5318
    Residual |  1.2062e+10       406  29709228.5   R-squared       =    0.0010
-------------+----------------------------------   Adj R-squared   =   -0.0015
       Total |  1.2074e+10       407  29664812.9   Root MSE        =    5450.6

------------------------------------------------------------------------------
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |   343.0828   548.2965     0.63   0.532    -734.7718    1420.937
       _cons |   5165.895   351.8358    14.68   0.000     4474.247    5857.542
------------------------------------------------------------------------------
```

Estimate impact with OLS: heterogeneity

Example: estimate impact for younger (less than and older than 24 y.o.)

```
.                  regress re78 treated if age > 24

      Source |       SS           df       MS      Number of obs   =       314
-------------+----------------------------------   F(1, 312)       =      3.79
       Model |   192959702         1   192959702   Prob > F        =    0.0526
    Residual |  1.5904e+10       312  50973253.4   R-squared       =    0.0120
-------------+----------------------------------   Adj R-squared   =    0.0088
       Total |  1.6097e+10       313  51426884.2   Root MSE        =    7139.6

------------------------------------------------------------------------------
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |   1593.373    818.946     1.95   0.053    -17.98248    3204.728
       _cons |   4991.653   524.9106     9.51   0.000     3958.841    6024.465
------------------------------------------------------------------------------
```

Non-experimental dataset (PSID)

- Now drop the observations for which the variable *randomization* equals 1.
- **What is now treatment and control group?**
  1. Treatment: individuals in the working-age population that applied to NSW and were admitted
  2. Control: individuals in the working-age population that applied to NSW and were NOT admitted + everybody else in the working-age population
- **Are they comparable? Is the counterfactual credible?**

Comparing balance for individual variables

The two groups are **not balanced at all**

```
.                          ttest age,   by(treat) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
-	2,490	34.8506	.209234	10.44076	34.44031	35.26089
Treated	297	24.62626	.3879837	6.686391	23.86271	25.38982
combined	2,787	33.76103	.200551	10.5875	33.36779	34.15428
diff		10.22434	.4408064		9.358228	11.09045

```
    diff = mean(-) - mean(Treated)                                 t =  23.1946
Ho: diff = 0                     Satterthwaite's degrees of freedom =  488.295

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Comparing balance for individual variables

The two groups are **not balanced at all**

```
.                          ttest educ,  by(treat) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
-	2,490	12.11687	.0617724	3.082435	11.99574	12.238
Treated	297	10.38047	.1054743	1.817712	10.1729	10.58805
combined	2,787	11.93183	.0572254	3.021046	11.81962	12.04403
diff		1.736396	.122232		1.496274	1.976518

```
    diff = mean(-) - mean(Treated)                              t =  14.2057
Ho: diff = 0                     Satterthwaite's degrees of freedom =  526.514

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Comparing balance for individual variables

The two groups are **not balanced at all**

```
.                          ttest black, by(treat) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
-	2,490	.2506024	.0086863	.433447	.2335692	.2676356
Treated	297	.8013468	.0231906	.3996597	.7557074	.8469862
combined	2,787	.3092931	.0087567	.4622852	.2921228	.3264635
diff		-.5507444	.024764		-.5994344	-.5020543

```
    diff = mean(-) - mean(Treated)                                t = -22.2397
Ho: diff = 0                         Satterthwaite's degrees of freedom =  383.983

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Comparing balance for individual variables

The two groups are **not balanced at all**

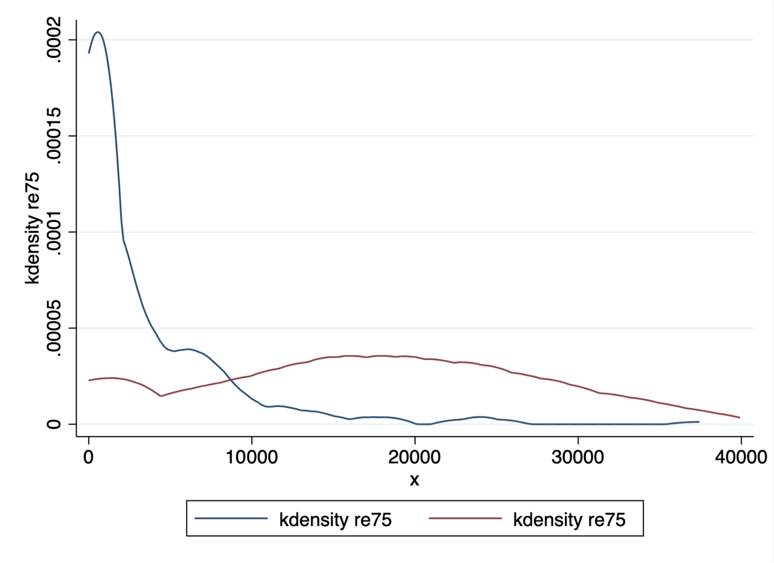
```
.                          ttest re75,  by(treat) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       - |   2,490    19063.34    272.4846    13596.95    18529.02    19597.66
 Treated |     297    3066.098    282.8697    4874.889    2509.407    3622.789
---------+--------------------------------------------------------------------
combined |   2,787    17358.57    262.5175    13858.84    16843.82    17873.32
---------+--------------------------------------------------------------------
    diff |            15997.24    392.7635                 15226.5    16767.98
------------------------------------------------------------------------------
    diff = mean(-) - mean(Treated)                                t =  40.7300
Ho: diff = 0                     Satterthwaite's degrees of freedom =  998.003

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Comparing balance for individual variables

The two groups are **not balanced at all**



Estimate difference with OLS: no controls

This is called **naive OLS estimator** – **why?**

```
.                          regress re78 treated

      Source |       SS           df       MS      Number of obs   =     2,787
-------------+----------------------------------   F(1, 2785)      =    290.90
       Model |  6.4390e+10         1  6.4390e+10   Prob > F        =    0.0000
    Residual |  6.1645e+11     2,785   221346575   R-squared       =    0.0946
-------------+----------------------------------   Adj R-squared   =    0.0942
       Total |  6.8084e+11     2,786   244379102   Root MSE        =     14878

------------------------------------------------------------------------------
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |  -15577.57   913.3285   -17.06   0.000    -17368.44    -13786.7
       _cons |   21553.92   298.1513    72.29   0.000      20969.3    22138.54
------------------------------------------------------------------------------
```

Estimate difference with OLS: controls

Controls are not very helpful in reducing bias in this case – **why?**

```
.                    regress re78 treated age age2 educ black hispanic nodegree

      Source |       SS           df       MS      Number of obs   =     2,787
-------------+----------------------------------   F(7, 2779)      =    121.50
       Model |  1.5954e+11         7  2.2791e+10   Prob > F        =    0.0000
    Residual |  5.2130e+11     2,779   187586428   R-squared       =    0.2343
-------------+----------------------------------   Adj R-squared   =    0.2324
       Total |  6.8084e+11     2,786   244379102   Root MSE        =     13696

------------------------------------------------------------------------------
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |  -8067.322    990.425    -8.15   0.000    -10009.37   -6125.279
         age |   1219.222    202.211     6.03   0.000     822.7232    1615.721
        age2 |  -14.21997   2.775828    -5.12   0.000    -19.66287   -8.777082
        educ |   1690.361   135.6443    12.46   0.000     1424.387    1956.334
       black |    -3204.3   655.1693    -4.89   0.000    -4488.968   -1919.632
    hispanic |   902.7708   1386.064     0.65   0.515    -1815.049    3620.591
    nodegree |   85.61254   856.3882     0.10   0.920    -1593.609    1764.834
       _cons |  -21850.51   3801.872    -5.75   0.000    -29305.29   -14395.73
------------------------------------------------------------------------------
```

What are we identifying in the case of IV?

IV deals with selection on unobservables

$$y_i = \beta + \alpha_i d_i + u_i$$

- **IV1 (homogeneity)**: $\alpha_i = \alpha$
- **IV2 (exclusion restriction)**: conditional on d, y is mean-independent of instrument z

$$E[y|d,z] = E[y|d] \text{ which implies } E[u|d,z] = E[u|d]$$

- **IV3 (relevance)**: there are at least two values of z (z^*, z^{**}) such that

$$P[d=1|z^*] \neq P[d=1|z^{**}]$$

Wald IV estimator

When the **instrument has only two values**, z^* and z^{**} (e.g. a dummy), we can derive IV estimator with a different procedure

- Consider the simplest case

$$y_i = \beta + \alpha d_i + u_i$$

- IV1 + IV2:

$$\begin{aligned} E(y_i|z_i = z^*) &= \beta + \alpha P(d_i = 1|z_i = z^*) + E(u_i) \\ E(y_i|z_i = z^{**}) &= \beta + \alpha P(d_i = 1|z_i = z^{**}) + E(u_i) \end{aligned}$$

Wald IV estimator

- By taking the difference:

$$E(y_i|z_i = z^*) - E(y_i|z_i = z^{**}) = \alpha[P(d_i = 1|z_i = z^*) - P(d_i = 1|z_i = z^{**})]$$

- Wald IV estimator

$$\alpha^{IV} = \frac{E[y_i|z_i = z^*] - E[y_i|z_i = z^{**}]}{P(d_i = 1|z_i = z^*) - P(d_i = 1|z_i = z^{**})}$$

- Notice the importance of IV3 in order to have a positive denominator (this is the rank condition in the IV estimator!)
- **Comparison with OLS?**

Identification of the true ATE

- **Identification of the true ATE** relies on:
  - homogeneity assumption (IV1)
- If IV1 doesn't hold, then in general IV identifies LATE

$$
\begin{aligned}
E(y_i|z_i = z^*) &= \beta + E[\alpha_i|z_i = z^*]P(d_i = 1|z_i = z^*) + E(u_i) \\
E(y_i|z_i = z^{**}) &= \beta + E[\alpha_i|z_i = z^{**}]P(d_i = 1|z_i = z^{**}) + E(u_i)
\end{aligned}
$$

- In first differences we obtain:

$$
\frac{E[y_i|z_i = z^*] - E[y_i|z_i = z^{**}]}{P(d_i = 1|z_i = z^*) - P(d_i = 1|z_i = z^{**})} = E[\alpha_i|z]
$$

- We need further assumptions!

An example: schooling as treatment

- Think about potential outcomes y_i

$$y_i = \begin{cases} y_{1i} \text{ if } d_i = 1 \text{ (complete schooling)} \\ y_{0i} \text{ if } d_i = 0 \text{ (drop-out)} \end{cases}$$

  - Allows writing: $y_i = y_{0i} + (y_{1i} - y_{0i})d_i$

- Instrument $z_i = \{0, 1\}$ with 2 values (simpler!) influences schooling - example: you have higher chance to go to school if you win in a lottery

$$d_i = \begin{cases} d_{1i} \text{ if } z_i = 1 \text{ (win lottery)} \\ d_{0i} \text{ if } z_i = 0 \text{ (lose lottery)} \end{cases}$$

  - Allows writing: $d_i = d_{0i} + (d_{1i} - d_{0i})z_i$

Assumptions

Potential outcomes can be indexed against schooling and z

$$y_i = \begin{cases} y_i(1,1) \text{ if } d_i = 1, z_i = 1 \\ y_i(1,0) \text{ if } d_i = 1, z_i = 0 \\ y_i(0,1) \text{ if } d_i = 0, z_i = 1 \\ y_i(0,0) \text{ if } d_i = 0, z_i = 0 \end{cases}$$

1. **Independence of instrument**

$$z_i \perp\!\!\!\perp \{y_{i0}, y_{i1}, d_{1i}, d_{0i}\} \tag{1}$$

2. **Relevance of instrument**

$$Cov(z, d) \neq 0 \tag{2}$$

3. **Monotonicity** ($d_{1i} - d_{0i}$ equals 1 or 0)

$$d_{1i} - d_{0i} \geq 0 \ \forall i \text{ (or viceversa)} \tag{3}$$

What is IV identifying?

Wald estimator

$$\frac{E[y_i|z_i = 1] - E[y_i|z_i = 0]}{P(d_i = 1|z_i = 1) - P(d_i = 1|z_i = 0)} = ?$$

- Start from the 1st term of the numerator:

$$\begin{aligned} E[y_i|z_i = 1] &= E[y_{0i} + (y_{1i} - y_{0i})d_i|z_i = 1] \\ &= E[y_{0i} + (y_{1i} - y_{0i})d_{1i}] \text{ by independence} \end{aligned}$$

- Same to the 2nd term, take difference and apply monotonicity:

$$\begin{aligned} E[y_i|z_i = 1] - E[y_i|z_i = 0] &= E[(y_{1i} - y_{0i})(d_{1i} - d_{0i})] \\ &= E[(y_{1i} - y_{0i})|d_{1i} > d_{0i}]P[d_{1i} > d_{0i}] \end{aligned}$$

- The denominator follows from the same derivation

$$E[d_i|z_i = 1] - E[d_i|z_i = 0] = E[d_{1i} > d_{0i}] = P[d_{1i} > d_{0i}]$$

LATE interpretation

Wald estimator as LATE

$$\frac{E[y_i|z_i = 1] - E[y_i|z_i = 0]}{P(d_i = 1|z_i = 1) - P(d_i = 1|z_i = 0)} = E[y_{1i} - y_{0i}|d_{1i} > d_{0i}]$$

- $d_{1i} > d_{0i} \Rightarrow$ individuals for whom the instrument changes the schooling decision (lottery winners)

	$d_{0i} = 0$	$d_{0i} = 1$
$d_{1i} = 0$	$\underbrace{y_i(0,1) - y_i(0,0) = 0}_{\text{Never taker}}$	$\underbrace{y_i(0,1) - y_i(1,0)}_{\textbf{Defier}}$
$d_{1i} = 1$	$\underbrace{y_i(1,1) - y_i(0,0)}_{\textbf{Complier}}$	$\underbrace{y_i(1,1) - y_i(1,0) = 0}_{\text{Always taker}}$

- **Different instruments will produce different LATEs!**

Imperfect compliance: ITT vs IV

Imagine out of 100 villages 50 are randomly receiving a treatment ($d = 1$) and 50 are controls ($d = 0$)

- **Imperfect compliance**
  - Some individuals in $d = 1$ do not receive treatment
  - $r_i = 1$ if received the treatment, 0 otherwise

1. OLS identifies what is called **Intent-to-Treat (ITT)**

$$y_i = X_i\beta + \alpha_i^{ITT} d_i + u_i$$

2. Use d as IV for r

$$\begin{aligned} y_i &= X_i\beta + \alpha_i r_i + u_i \\ r_i &= X_i\beta + d_i\gamma + v_i \end{aligned}$$

APPLICATION: back to Lalonde (1986) dataset

- As before we make use of the observations from PSID – drop the observations for which the variable *randomization* equals 1.
- **How can we apply IV to this setting?**
  1. We need to find an **instrument** for the variable treated
     - Use the dummy variable "married" (equal to 1 if the individual is married and equal to 0 otherwise)
     - **Relevance**: correlated with *treated* indicator
     - **Exclusion restriction**: not correlated with unobservable determinants of earnings (*re*78)
  2. **Is this a good instrument?**

IV without controls

First stage 2SLS estimates

```
.                          ivreg re78 (treated = married), first

First-stage regressions
```

Source	SS	df	MS
Model	74.670529	1	74.670529
Residual	190.67931	2,785	.068466538
Total	265.349839	2,786	.09524402

Number of obs	=	2,787
F(1, 2785)	=	1090.61
Prob > F	=	0.0000
R-squared	=	0.2814
Adj R-squared	=	0.2811
Root MSE	=	.26166

treated	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
married	-.4032069	.0122093	-33.02	0.000	-.4271472	-.3792666
_cons	.4258621	.0108649	39.20	0.000	.404558	.4471661

IV without controls

First stage   **2SLS estimates**

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	3.9134e+10	1	3.9134e+10
Residual	6.4171e+11	2,785	230414981
Total	6.8084e+11	2,786	244379102

Number of obs	=	2,787
F(1, 2785)	=	207.98
Prob > F	=	0.0000
R-squared	=	0.0575
Adj R-squared	=	0.0571
Root MSE	=	15179

re78	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
treated	-25333.5	1756.632	-14.42	0.000	-28777.93	-21889.07
_cons	22593.57	343.1003	65.85	0.000	21920.82	23266.33

Instrumented: treated
Instruments: married

IV with controls - how to interpret?

First stage   2SLS estimates

```
.                        ivreg re78 age educ black hisp nodeg re75 (treated = married), first

First-stage regressions
```

Source	SS	df	MS
Model	104.707917	7	14.9582739
Residual	160.641922	2,779	.057805657
Total	265.349839	2,786	.09524402

Number of obs	=	2,787
F(7, 2779)	=	258.77
Prob > F	=	0.0000
R-squared	=	0.3946
Adj R-squared	=	0.3931
Root MSE	=	.24043

treated	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0037528	.0004841	-7.75	0.000	-.0047021	-.0028036
educ	.0101419	.0024248	4.18	0.000	.0053874	.0148965
black	.1222855	.0113776	10.75	0.000	.0999762	.1445949
hispanic	.1475669	.0241656	6.11	0.000	.1001826	.1949512
nodegree	.1323333	.0148324	8.92	0.000	.1032497	.161417
re75	-2.66e-06	3.84e-07	-6.91	0.000	-3.41e-06	-1.90e-06
married	-.284564	.0125818	-22.62	0.000	-.3092346	-.2598933
_cons	.2937148	.0394143	7.45	0.000	.2164306	.3709991

IV with controls - how to interpret?

First stage   2SLS estimates

Instrumental variables (2SLS) regression

Source	SS	df	MS
Model	3.8893e+11	7	5.5562e+10
Residual	2.9191e+11	2,779	105040761
Total	6.8084e+11	2,786	244379102

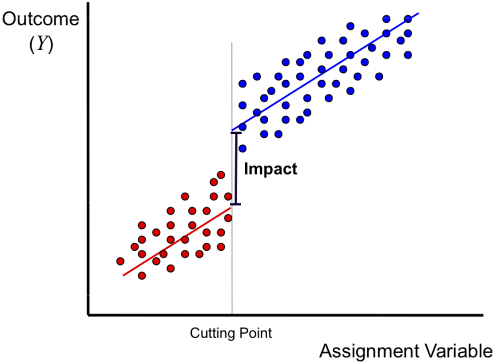
Number of obs	=	2,787
F(7, 2779)	=	535.47
Prob > F	=	0.0000
R-squared	=	0.5713
Adj R-squared	=	0.5702
Root MSE	=	10249

re78	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
treated	-7112.459	1884.767	-3.77	0.000	-10808.14	-3416.775
age	-76.54291	23.40888	-3.27	0.001	-122.4435	-30.64236
educ	730.5017	106.9618	6.83	0.000	520.7691	940.2342
black	147.2899	575.544	0.26	0.798	-981.2471	1275.827
hispanic	2746.651	1077.502	2.55	0.011	633.8649	4859.437
nodegree	1332.688	706.7127	1.89	0.059	-53.04686	2718.423
re75	.7638929	.0179238	42.62	0.000	.7287475	.7990382
_cons	639.5256	1644.599	0.39	0.697	-2585.234	3864.285

Instrumented: treated
Instruments: age educ black hispanic nodegree re75 married

Another LATE estimator ⇒ regression discontinuity (RD)

Probability of treatment changes **discontinuously** with some **observable continuous variable** z (*assignment or forcing variable*)

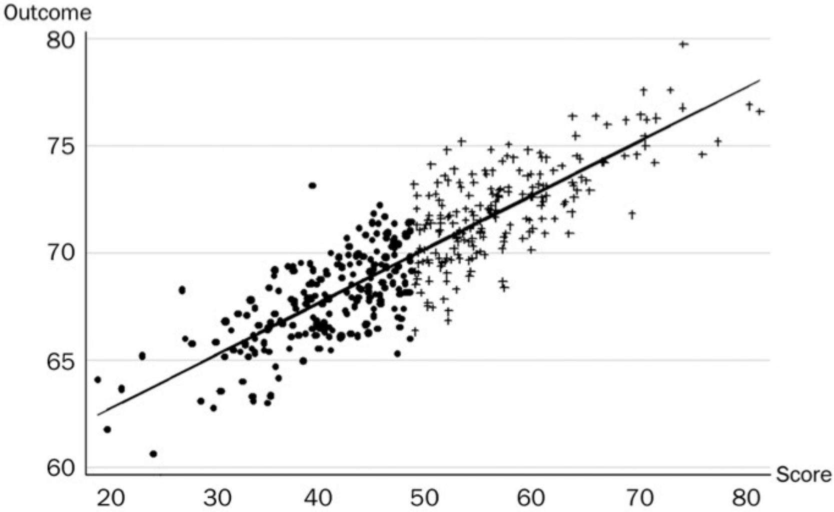


Examples:

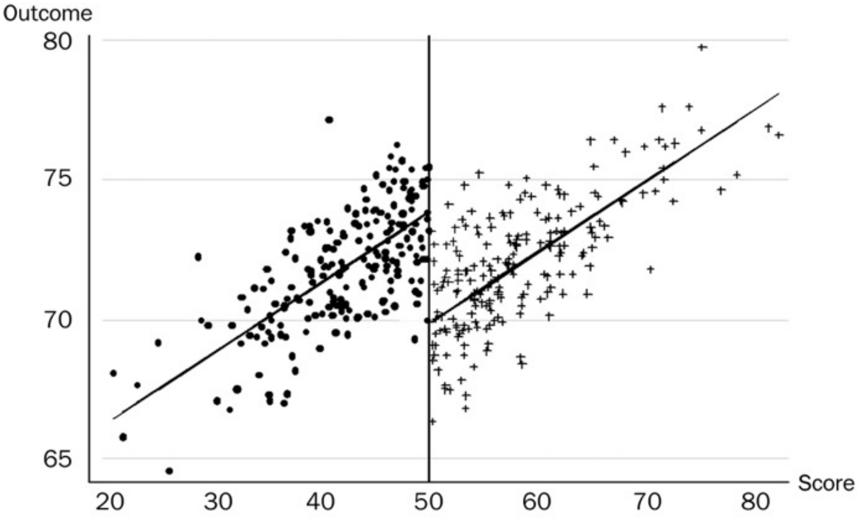
- Students receive a scholarship if GPA is ≥ 3.0
- Individuals eligible for a loan if they own < 0.5 acres of land
- Legislators are elected if they obtain > 50% of votes

Main idea: on either sides of the cut-off, individuals are very similar, but treatment status differs

RD in practice: pre-programme or unaffected variables



RD in practice: post-programme



RD setting

$$y_i = \beta_i + \alpha_i d_i + u_i$$

Assumptions needed for identification:

1. **Discontinuity**: d is a function of z discontinuous at $z = z^*$

$$\lim_{z \to z^{*-}} P(d = 1|z) \neq \lim_{z \to z^{*+}} P(d = 1|z)$$

2. **Smoothness**: $E[\beta_i|z]$ and $E[\alpha_i|z]$ are continuous at $z = z^*$

$$\lim_{z \to z^{*-}} E[\beta_i|z] = \lim_{z \to z^{*+}} E[\beta_i|z]; \qquad \lim_{z \to z^{*-}} E[\alpha_i|z] = \lim_{z \to z^{*+}} E[\alpha_i|z]$$

3. **Local randomization**: α_i independent from d in the neighbourhood of z^*

RD setting

- Potential outcomes $E[y_{i0}|z]$ and $E[y_{i1}|z]$ are continuous at $z = z^*$
- For each value of $z \Rightarrow$ observe either $E[y_{i0}|z]$ OR $E[y_{i1}|z]$

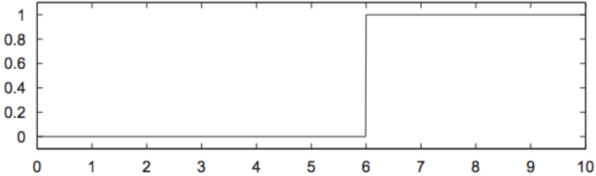


Fig. 1. Assignment probabilities (SRD).

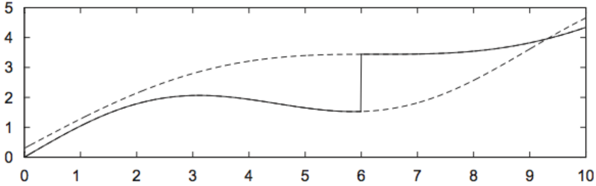


Fig. 2. Potential and observed outcome regression functions.

Identification

1. Define $p(z^*) \equiv P(d_i = 1|z = z^*)$ and compute $E(y_i|z^*)$:

$$\begin{aligned} E(y_i|z^*) &= E(\beta_i|z^*) + p(z^*) \cdot E(\alpha_i|d = 1, z^*) \\ &= E(\beta_i|z^*) + p(z^*) \cdot E(\alpha_i|z^*) \end{aligned}$$

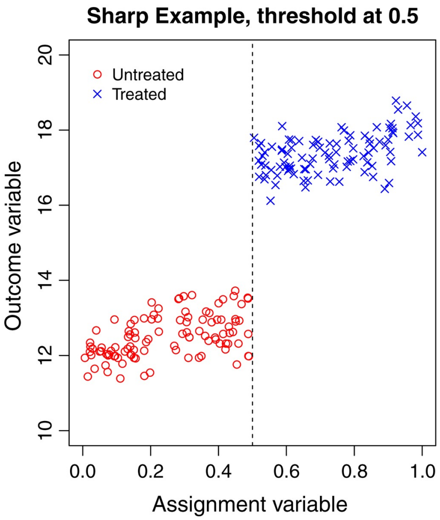
2. Take **difference in limits** around the cut-off

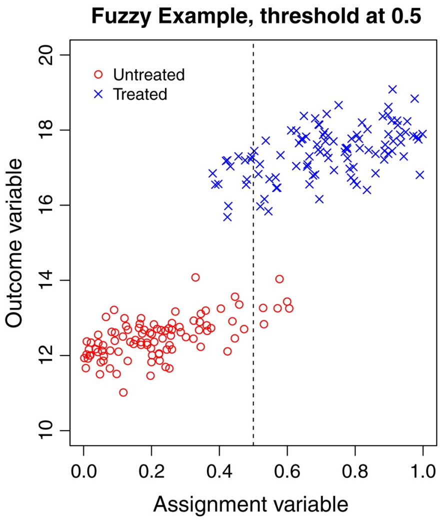
$$\lim_{z \to z^{*+}} E[y_i|z] - \lim_{z \to z^{*-}} E[y_i|z] = E[\alpha_i|z^*]\left[\lim_{z \to z^{*+}} p(z) - \lim_{z \to z^{*-}} p(z)\right]$$

   - $E[\alpha|z^*]$ is a **LATE** $\Rightarrow$
     - Average effect for those at the discontinuity ($z = z^*$)
     - We do not learn about α_i away from the discontinuity

3. Final formula depends of the features of the discontinuity

Sharp versus fuzzy designs





RD: sharp versus fuzzy designs

1. **Fuzzy RD**: $p(z)$ is in between 0 and 1

$$\alpha^{RD,FUZZY}(z^*) = \frac{\lim_{z \to z^{*+}} E[y_i|z] - \lim_{z \to z^{*-}} E[y_i|z]}{\lim_{z \to z^{*+}} p(z) - \lim_{z \to z^{*-}} p(z)} \tag{4}$$

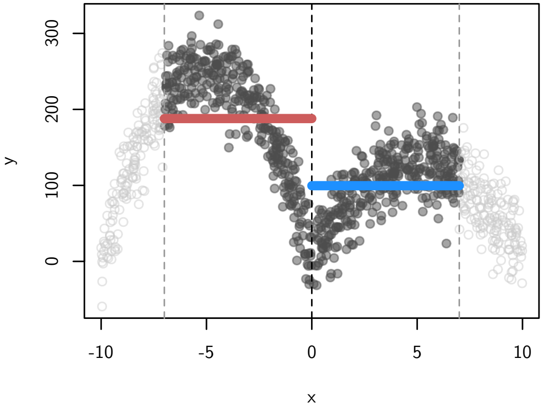
2. **Sharp RD**: $p(z)$ is either 0 or 1 in different sides of the cut-off → denominator of equation (4) is equal to 1

$$\alpha^{RD,SHARP}(z^*) = \lim_{z \to z^{*+}} E[y_i|z] - \lim_{z \to z^{*-}} E[y_i|z] \tag{5}$$

Non-parametric estimation in sharp RD

Estimator: sample correspondent of equation (5) restricting the sample to **bandwidth** $z^* \pm \Delta$

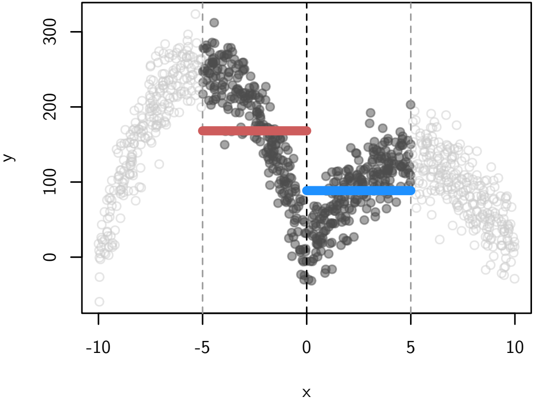
- **Efficiency-bias trade-off**: $\downarrow \Delta \Rightarrow \uparrow$ similarity of individuals around the discontinuity $\downarrow$ precision (less observations)



Non-parametric estimation in sharp RD

Estimator: sample correspondent of equation (5) restricting the sample to **bandwidth** $z^* \pm \Delta$

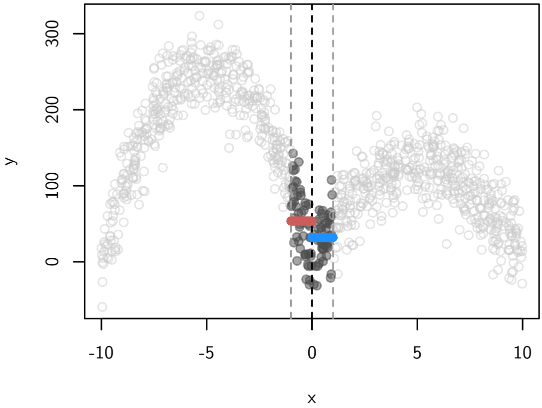
- **Efficiency-bias trade-off**: $\downarrow \Delta \Rightarrow \uparrow$ similarity of individuals around the discontinuity $\downarrow$ precision (less observations)



Non-parametric estimation in sharp RD

Estimator: sample correspondent of equation (5) restricting the sample to **bandwidth** $z^* \pm \Delta$

- **Efficiency-bias trade-off**: $\downarrow \Delta \Rightarrow \uparrow$ similarity of individuals around the discontinuity $\downarrow$ precision (less observations)



Parametric estimation in sharp RD

Estimator: explicitly estimate the conditional mean of y as function of z and look at the jump at the cut-off

- Some examples using OLS to $E[y_i|z]$:
  1. $f(z)$ is **linear** $\Rightarrow$ equivalent to local conditional means

  $$y_i = \beta + \alpha d_i + \gamma z_i + \epsilon_i$$

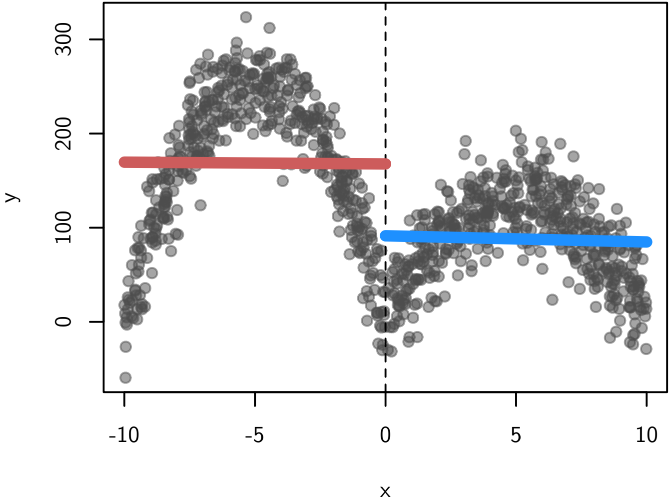
  2. $f(z)$ **behaves differently on either side of cut-off**

  $$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$

     - For correct interpretation of α (see interaction terms) $\Rightarrow$ make sure z_i is discontinuous at 0 or use the transform $\tilde{z}_i = z_i - z^*$
     - More flexible forms $\Rightarrow$ adds z (and interactions) with powers higher than 1

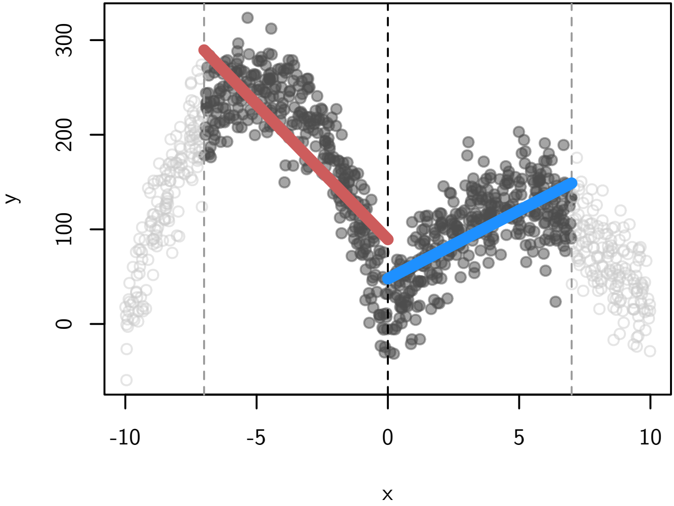
Example: allow slopes and intercepts to change

$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$



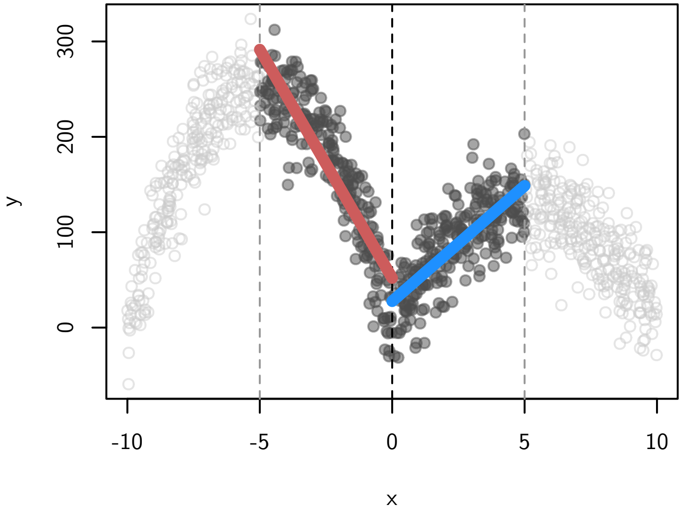
Example: allow slopes and intercepts to change

$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$



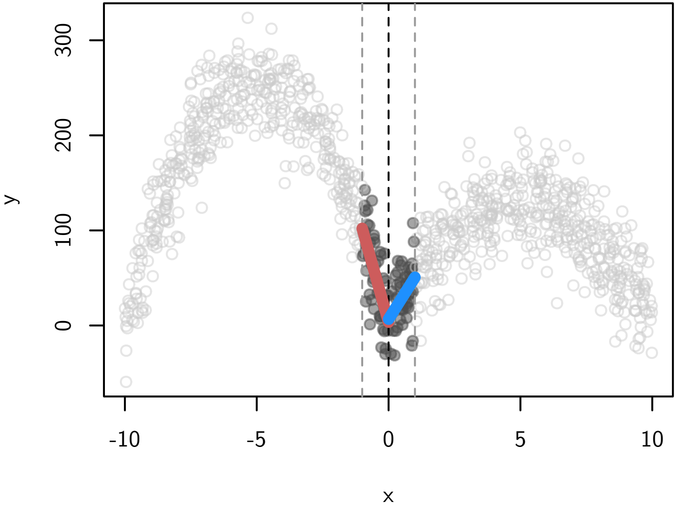
Example: allow slopes and intercepts to change

$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$



Example: allow slopes and intercepts to change

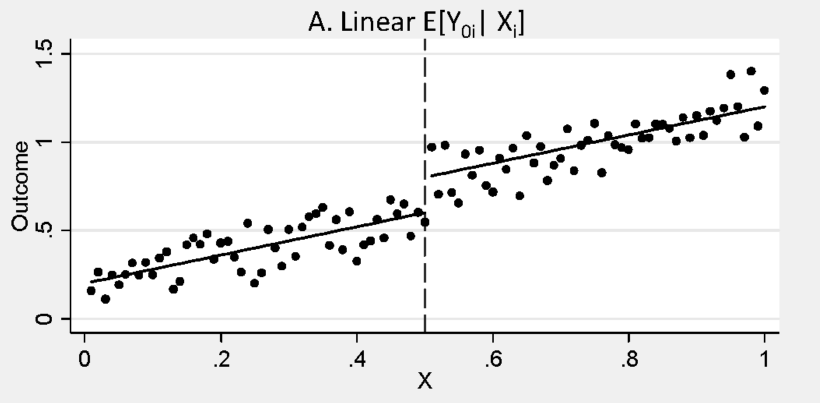
$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$



Getting the right functional form

Functions can be different: what is the right assumption?

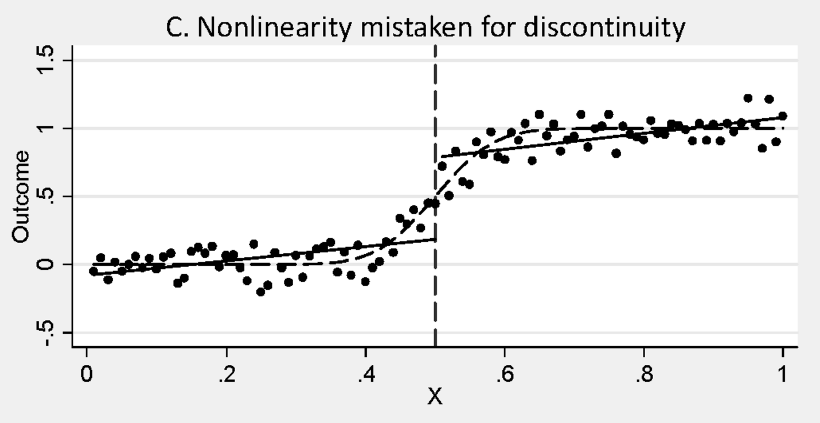
$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$



Getting the right functional form

Functions can be different: what is the right assumption?

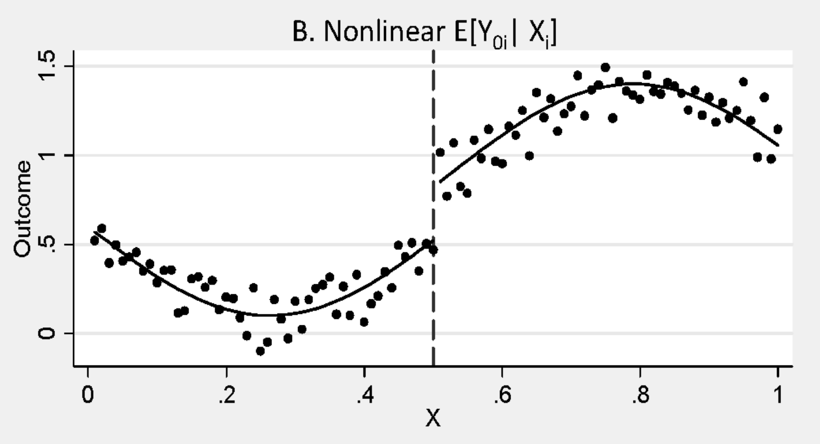
$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \mu_1 d_i z_i + \epsilon_i$$



RD: getting the right functional form

Functions can be different: include higher-degree interactions

$$y_i = \beta + \alpha d_i + \gamma_1 z_i + \gamma_2 z_i^2 + \mu_1 d_i z_i + \mu_2 d_i z_i^2 + \epsilon_i$$

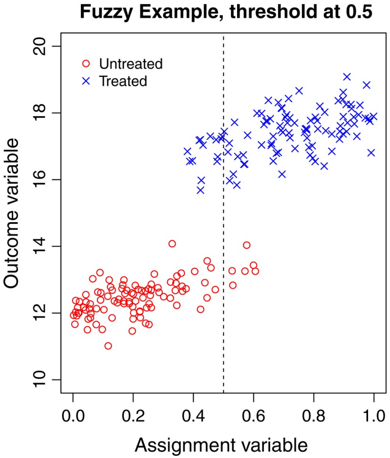


Estimation with fuzzy RD

1. **Non-parametric:** sample correspondents of equation (4)
2. **Parametric:** apply Wald estimator (or 2SLS) to identify LATE

$$\frac{E[y_i|z^*] - E[y_i|z^{**}]}{P(d_i = 1|z^*) - P(d_i = 1|z^{**})}$$

- z is a **perfect IV**
  - uncorrelated with ϵ_i (exclusion restriction)
  - correlated with d_i (relevance)



APPLICATION: Lemieux and Milligan (2004)

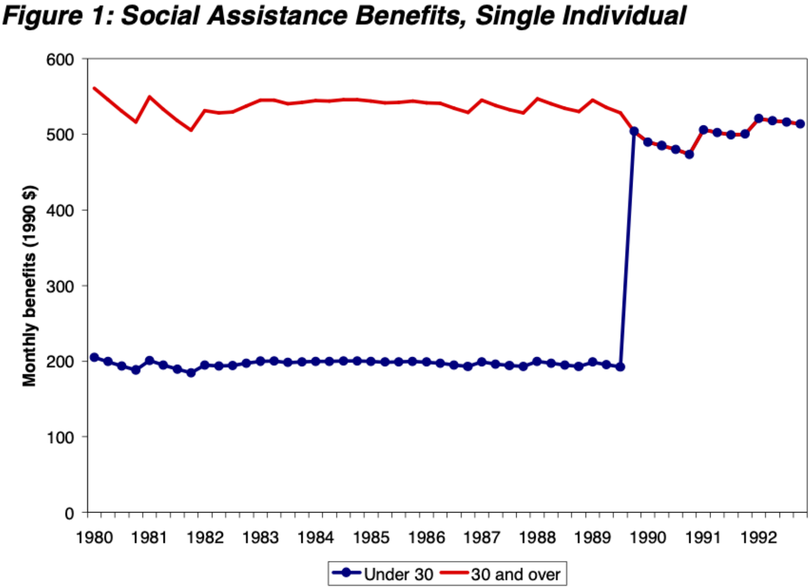
How the provision of social assistance affects labour supply?

- **DISCONTINUOUS change in benefits** in Canada

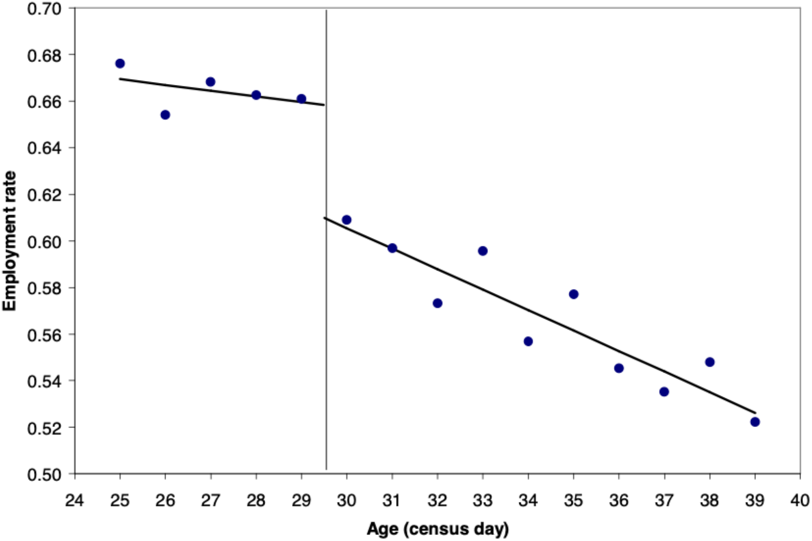
We examine the incentive effects of transfer programs using a unique policy episode. Prior to 1989, social assistance recipients without children in Quebec who were under age 30 received benefits 60 percent lower than recipients older than 30. We use this sharp discontinuity in policy to estimate the effects of social assistance on various labour market outcomes and on living arrangements using a regression discontinuity approach. We find strong evidence that more generous social assistance benefits reduce employment, and more suggestive evidence that they affect marital status and living arrangements. The regression discontinuity estimates exhibit little sensitivity to the degree of flexibility in the specification, and perform very well when we control for unobserved heterogeneity using a first difference specification. Finally, we show that commonly used difference-in-difference estimators may perform poorly when control groups are inappropriately chosen.

- **SHARP design** based on *age* at $age^* = 30$

The origin of the discontinuity



RD estimates: $E[y|z]$ is assumed linear with different slopes and intercepts



RD estimates

Specification for age	Empl. rate last year		Empl. Rate at census		Difference in empl. rate		Weekly hours	
Mean of the dependent variable								
	0.562		0.618		0.056		24.39	
Regression discontinuity estimates								
Linear	-0.045	***	-0.041	***	-0.029	**	-1.45	**
	(0.012)		(0.012)		(0.011)		(0.54)	
Quadratic	-0.048	***	-0.051	***	-0.031	**	-1.75	**
	(0.013)		(0.012)		(0.012)		(0.61)	
Cubic	-0.043	**	-0.048	***	-0.030	**	-1.47	*
	(0.018)		(0.014)		(0.013)		(0.70)	
Linear spline	-0.047	***	-0.049	***	-0.032	**	-1.72	***
	(0.013)		(0.011)		(0.013)		(0.55)	
Quadratic spline	-0.038		-0.056	**	-0.035	*	-1.66	
	(0.024)		(0.018)		(0.016)		(0.94)	

Are assumption valid? Check continuity