# RDD and diff-in-diff

Silje Synnøve Lyder Hermansen

03-12-2019

# Regression discontinuity design (RDD)

# Basic assumption

**RDD assumes a running variable ($x$) with a cut point ($c$) beyond which treatment is assigned ($D$).**

$$D_i = \begin{cases} 1 & if \quad x_i \quad \geqslant c \\ 0 & if \quad x_i \quad < c \end{cases} \tag{1}$$

# Distinction

**It has a flavor of logit or propensity scores, but there are some differences:**

- ~~**logit**~~ : $x$ is not latent and we know the cutpoint: Both are observed and included as a *predictors*.
- ~~**matching**~~ : we have no control/treatment group. However, we assume that units on either side of the treatment are increasingly similar as their $x$ is similar.

# Examples

**Administrative data are perfect:** You have some rule that kicks in at a specific threshold for otherwise almost identical observations.

- school test scores on school admission, restrictions on class size
- legal drinking age on alcohol related deaths
- election of candidates in close races

# Two ways of understanding RDD

- **Individuals close to the threshold are interchangeable**

# Two ways of understanding RDD

- **Individuals close to the threshold are interchangeable**

$\rightarrow$ in a small window, you have a treatment and a control group.

# Two ways of understanding RDD

- **Individuals close to the threshold are interchangeable**

$\rightarrow$ in a small window, you have a treatment and a control group.

- $x$ **is a bottleneck:** the relationship between $D$ and $Y$ is confounded by $x$, but all other confounders only influence $Y$ through $x$.

# Two ways of understanding RDD

- **Individuals close to the threshold are interchangeable**

$\rightarrow$ in a small window, you have a treatment and a control group.

- $x$ **is a bottleneck:** the relationship between $D$ and $Y$ is confounded by $x$, but all other confounders only influence $Y$ through $x$.

$\rightarrow$ conditioning on $x$ is sufficient to isolate the causal effect.

# Two ways of understanding RDD

- **Individuals close to the threshold are interchangeable**

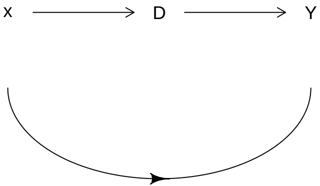
$\rightarrow$ in a small window, you have a treatment and a control group.

- $x$ **is a bottleneck:** the relationship between $D$ and $Y$ is confounded by $x$, but all other confounders only influence $Y$ through $x$.

$\rightarrow$ conditioning on $x$ is sufficient to isolate the causal effect.

# Two ways of understanding RDD

**X is a confounder**
**...so we only control for X**

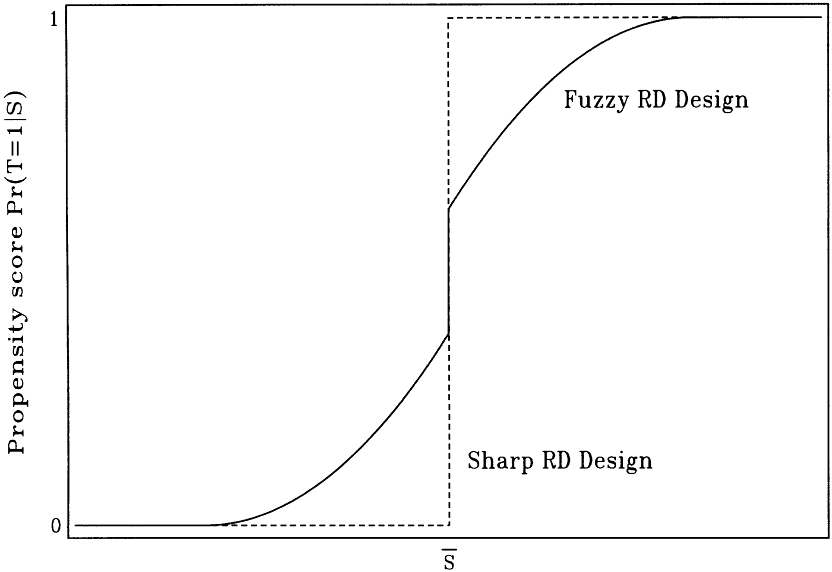


# Two designs

**We distinguish between two designs depending on how probable the treatment is:**

- **sharp** RD: assignment is deterministic
- **fuzzy** RD: assignment is probabilistic

# A visual representation



# The basic model

# The basic regression

**We assume the relationship between x and y is linear and the treatment is deterministic**

# The basic regression

**We assume the relationship between x and y is linear and the treatment is deterministic**

$$y_i = \alpha + \rho \times D_i + \gamma \times x_1 + e_i \tag{2}$$

# The basic regression

**We assume the relationship between x and y is linear and the treatment is deterministic**

$$y_i = \alpha + \rho \times D_i + \gamma \times x_1 + e_i \tag{2}$$

$\Rightarrow$ *The treatment is reported by* $\rho$

# The basic regression

**We assume the relationship between x and y is linear and the treatment is deterministic**

$$y_i = \alpha + \rho \times D_i + \gamma \times x_1 + e_i \tag{2}$$

$\Rightarrow$ *The treatment is reported by* $\rho$

# Do it in R

**Check out alcohol related deaths ($y$) as a function of legal drinking age ($D$)**

```
load("df.Rda")
mod1 <- lm(alcohol ~ over21 + age,
           df)
```

# Is that all?

**This is true on two conditions**

# Is that all?

**This is true on two conditions**

1. **the continuity assumption**: $x$ must have a continuous effect on $y$

# Is that all?

**This is true on two conditions**

1. **the continuity assumption**: $x$ must have a continuous effect on $y$
2. **no omited variable bias**:$x$ must capture all influence on $D$.

# The continuity assumption

# The continuity assumption

**Sometimes we may pick up a smooth non-linear change by dummy coding**

# The continuity assumption

**Sometimes we may pick up a smooth non-linear change by dummy coding**

. . . that's not a regression discontinuity.

# Ensuring linear effect

**We can obtain a linear effect in two ways:**

# Ensuring linear effect

**We can obtain a linear effect in two ways:**

- recode the $x$

# Ensuring linear effect

**We can obtain a linear effect in two ways:**

- recode the $x$ $\rightarrow$ compare with the recoding of $y$ in GLMs.

# Ensuring linear effect

**We can obtain a linear effect in two ways:**

- recode the $x \rightarrow$ compare with the recoding of $y$ in GLMs.
- consider a sufficiently small window

# Recode the x

**We can create a curvilinear effect of x using polynomials** (e.g.:)

# Recode the x

**We can create a curvilinear effect of x using polynomials** (e.g.:)

$$y_i = \alpha + \rho D_i + \gamma_1 x_i + \gamma_2 x_i^2 \tag{3}$$

# Recode the x

**We can create a curvilinear effect of x using polynomials** (e.g.:)

$$y_i = \alpha + \rho D_i + \gamma_1 x_i + \gamma_2 x_i^2 \tag{3}$$

```
df$age2 <- df$age^2
mod2 <- lm(alcohol ~ over21 + age + age2,
           df)
```

# Recode the x: symmetric effect

**We can create a curvilinear effect of x using polynomials** (e.g.:)

$$y_i = \alpha + \rho D_i + \gamma_1 x_i + \gamma_2 x_i^2 \tag{4}$$

```
df$age2 <- df$age^2
mod2 <- lm(alcohol ~ over21 + age + age2,
           df)
```

⇒ *Here, $x$ has a symmetrical effect on both sides of the treatment.*

# Recode the x: assymetric effect

**We can assume x has different effects on each side of the treatment**

$$y_i = \alpha + \rho D_i + \gamma x_i + \delta x_i D_i \tag{5}$$

# Recode the x: assymetric effect

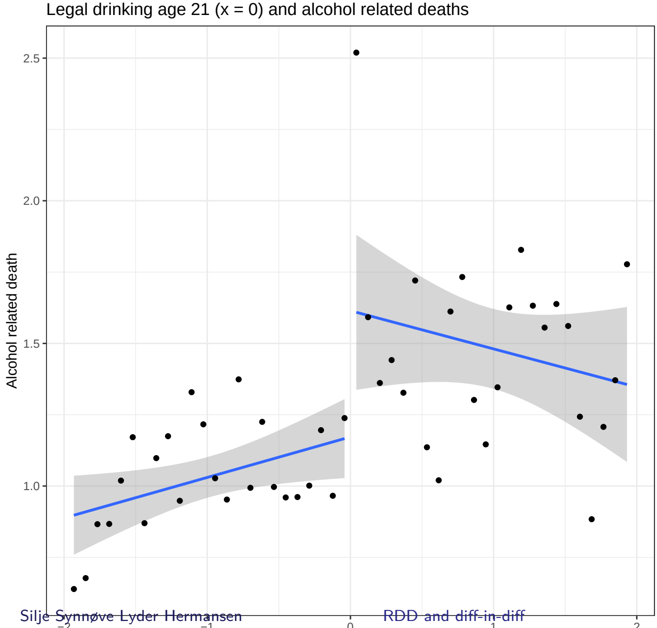
**We can assume x has different effects on each side of the treatment**

$$y_i = \alpha + \rho D_i + \gamma x_i + \delta x_i D_i \tag{5}$$

$\Rightarrow$ *we center the x on the cutpoint* $(x_i - c) \rightarrow \rho$ *still reports the change at the cutpoint.*

```
mod3 <- lm(alcohol ~ over21 * age,
          df)
```

# Do it in R


Legal drinking age 21 (x = 0) and alcohol related deaths
Alcohol related death
2.5
2.0
1.5
1.0
0
1
2

# Extrapolation

**We do this to estimate the effect at the cutpoint ($\rho$)**

# Extrapolation

**We do this to estimate the effect at the cutpoint ($\rho$)**

- but we can also extrapolate $y$ beyond the cutpoint with $x$: $\rho + \delta(x - c)$

# Recode the x

- recoding the $x$ is a **the parametric approach**.

# Recode the x

- recoding the $x$ is a **the parametric approach**.
- subsetting the data to tweak the window around the cutpoint is a **non-parametric approach.**

# Bandwidth : the idea

**If the span of $x$ around $c$ is sufficiently small, there is no problem with non-linearity**

# Bandwidth : the idea

**If the span of $x$ around $c$ is sufficiently small, there is no problem with non-linearity**

- There's a tradeoff between linearity and statistical power (we need sufficient N).

# Bandwidth : how do we choose it?

**We try out different bandwidths**

# Bandwidth : how do we choose it?

**We try out different bandwidths**

- We can do it by hand

# Bandwidth : how do we choose it?

**We try out different bandwidths**

- We can do it by hand

```
mod5 <- lm(alcohol ~ over21 * age,
           df[df$agecell > 20 & df$agecell < 22,])
```

# Bandwidth : how do we choose it?

**We try out different bandwidths**

- We can do it by hand

```
mod5 <- lm(alcohol ~ over21 * age,
           df[df$agecell > 20 & df$agecell < 22,])
```

$\Rightarrow$ *When you narrow down, do you get a weaker or stronger effect?*

# Bandwidth : how do we choose it?

**We try out different bandwidths**

- We can do it by hand

# Bandwidth : how do we choose it?

**We try out different bandwidths**

- We can do it by hand
- . . . or we can make an algorithm do it:
  - run a local weighted regression line
  - bandwidth is estimated accordingly

$\Rightarrow$ *the point is to show robustness, not p-hack!*

# Omited variable bias

# Fuzzy RD

# Fuzzy RD

**Often the $D$ increases the probability of a treatment, but we don't know!**

$\Rightarrow$ *This is a Instrumental Variable approach (more on Thursday)*

# In brief

$x$ **has a unique effect on** $D$**.** I'm interested in the effect of $\bar{x}$ on $y$, but $x$ is completely endogeneous:

# In brief

$x$ **has a unique effect on** $D$**.** I'm interested in the effect of $\bar{x}$ on $y$, but $x$ is completely endogeneous:

- $y = \alpha + \beta_1 \bar{x} + \beta_2 x + e$

# In brief

$x$ **has a unique effect on** $D$**.** I'm interested in the effect of $\bar{x}$ on $y$, but $x$ is completely endogeneous:

- $y = \alpha + \beta_1 \bar{x} + \beta_2 x + e$

**I use treatment as an instrument. We do this in two steps**

# In brief

$x$ **has a unique effect on** $D$**.** I'm interested in the effect of $\bar{x}$ on $y$, but $x$ is completely endogeneous:

- $y = \alpha + \beta_1 \bar{x} + \beta_2 x + e$

**I use treatment as an instrument. We do this in two steps**

- step 1: $\bar{x} = \alpha_1 + \phi D + \beta_1 x + e_1$

# In brief

$x$ **has a unique effect on** $D$**.** I'm interested in the effect of $\bar{x}$ on $y$, but $x$ is completely endogeneous:

- $y = \alpha + \beta_1 \bar{x} + \beta_2 x + e$

**I use treatment as an instrument. We do this in two steps**

- step 1: $\bar{x} = \alpha_1 + \phi D + \beta_1 x + e_1$
- step 2: $y = \alpha_2 + \gamma \tilde{x} + \beta_2 x + e_2$

# In brief

$x$ **has a unique effect on** $D$**.** I'm interested in the effect of $\bar{x}$ on $y$, but $x$ is completely endogeneous:

- $y = \alpha + \beta_1 \bar{x} + \beta_2 x + e$

**I use treatment as an instrument. We do this in two steps**

- step 1: $\bar{x} = \alpha_1 + \phi D + \beta_1 x + e_1$
- step 2: $y = \alpha_2 + \gamma \tilde{x} + \beta_2 x + e_2$

$\Rightarrow \gamma$ is the causal effect of D in a fuzzy design.

# Differences-in-differences

# Definition: Comparing two differences

# Definition: Differences-in-differences

**Treatment and control groups may differ in many ways (they are not randomly assigned)**

# Definition: Differences-in-differences

**Treatment and control groups may differ in many ways (they are not randomly assigned)**

- Pre-treatment: They move in parallel

# Definition: Differences-in-differences

**Treatment and control groups may differ in many ways (they are not randomly assigned)**

- Pre-treatment: They move in parallel
- Post-treatment: They diverge

# Definition: Differences-in-differences

**Treatment and control groups may differ in many ways (they are not randomly assigned)**

- Pre-treatment: They move in parallel
- Post-treatment: They diverge

$\Rightarrow$ *Treatment effect is that difference*

# What differences?

**Diff-in-diff is based on two comparisons**

# What differences?

**Diff-in-diff is based on two comparisons**

- the difference pre- and post treatement *within* each unit

# What differences?

**Diff-in-diff is based on two comparisons**

- the difference pre- and post treatement *within* each unit
- the difference *between* the treatment and control groups

# What differences?

**Diff-in-diff is based on two comparisons**

- the difference pre- and post treatement *within* each unit
- the difference *between* the treatment and control groups

$\Rightarrow$ *based on panel data (units are observed several times).*

# Example: States' monetary policy and number of banks

**Take the differences between number of banks in two districts in Missisippi**

# Example: States' monetary policy and number of banks

**Take the differences between number of banks in two districts in Missisippi**

- Treatment: District 6 provided money to banks, while district 8 did not.

# Example: States' monetary policy and number of banks

**Take the differences between number of banks in two districts in Missisippi**

- Treatment: District 6 provided money to banks, while district 8 did not.
- After a year district 6 had 121 banks, while district 8 had 132

# Example: States' monetary policy and number of banks

**Take the differences between number of banks in two districts in Missisippi**

- Treatment: District 6 provided money to banks, while district 8 did not.
- After a year district 6 had 121 banks, while district 8 had 132

$\Rightarrow$ *What was the treatment effect?*

# Example: States' monetary policy and number of banks

**The two districts started out differently**

# Example: States' monetary policy and number of banks

**The two districts started out differently**

- within-unit difference: Number of banks before and after the crisis in each district.

District 6: \$121 - 135 = \$ −14; District 8: \$132 - 165 = \$−33

# Example: States' monetary policy and number of banks

**The two districts started out differently**

- within-unit difference: Number of banks before and after the crisis in each district.

District 6: \$121 - 135 = \$ −14; District 8: \$132 - 165 = \$−33

- between-unit differences: take the difference between the two.

$-14 - (-33) = 19$

# Example: States' monetary policy and number of banks

**The two districts started out differently**

- within-unit difference: Number of banks before and after the crisis in each district.

District 6: \$121 - 135 = \$ −14; District 8: \$132 - 165 = \$−33

- between-unit differences: take the difference between the two.

$-14 - (-33) = 19$

$\Rightarrow$ *Basically a 2-by-2 table*

# How to do it?

# Interaction effects

**In a regression, these differences are represented by an interaction term between two dummies**

# Interaction effects

**In a regression, these differences are represented by an interaction term between two dummies**

$$y_i = \alpha + \beta_1 T_i + \beta_2 P + \beta_3 T_i P_i \tag{6}$$

# Interaction effects

**In a regression, these differences are represented by an interaction term between two dummies**

$$y_i = \alpha + \beta_1 T_i + \beta_2 P + \beta_3 T_i P_i \quad (6)$$

- $P$ represents post-treatment effect: differences *within* units

# Interaction effects

**In a regression, these differences are represented by an interaction term between two dummies**

$$y_i = \alpha + \beta_1 T_i + \beta_2 P + \beta_3 T_i P_i \tag{6}$$

- $P$ represents post-treatment effect: differences *within* units
- $T$ represents the treatment group: differences *between* units

# Interaction effects

**In a regression, these differences are represented by an interaction term between two dummies**

$$y_i = \alpha + \beta_1 T_i + \beta_2 P + \beta_3 T_i P_i \tag{6}$$

- $P$ represents post-treatment effect: differences *within* units
- $T$ represents the treatment group: differences *between* units
- $\beta_3$ is the causal effect

# Data

**Data requirements**

- Requires panel data

# Data

**Data requirements**

- Requires panel data → which means correcting the standard errors.
- Common panel types: state-year/administrative unit-time period; people over time ...

⇒ *we want to know the trend before and after the break*

# Another example: drinking age and death

# Another example: drinking age and death

**Does the legal drinking age has an effect on death rates among the young?**

- $y$ is number of deaths per 100 000
- $P$ is post-treatment dummy
- $T$ is dummies for states
- *trend* is year dummies

# Another example: step 1 → calculate differences

The authors have two tricks:

- Hardcode the interaction effect (dummy before/after treatment)
- They remove the intercept to retain all dummies

```
load("df2.Rda")
##with intercept
mod <- lm(mrate ~ legal +
            state +
            year_fct,
          df)
##without intercept; with all dummies
mod <- lm(mrate ~ 0 +
            legal +
            state +
            year_fct,
          df)
```

# Another example: step 2 $\rightarrow$ calculate errors

Calculate robust standard errors:

```
library(clubSandwich)
```

```
## Registered S3 method overwritten by 'clubSandwich':
##   method    from
##   bread.mlm sandwich
```

```
vcov <- vcovCR(mod, cluster = df[["state"]],
               type = "CR2")
robust <- coef_test(mod, vcov = vcov)$SE
```

# Another example: step 3 → interpretation

Display the results and interpret:

```r
library(stargazer)
stargazer(mod, se = robust,
                    omit = "state|year",
                    type = "html",
                    out = "regtable.html")
```

# Another example: step 3 $\rightarrow$ interpretation

Table 1: Death rates among young as a function of legal drinking age

| | *Dependent variable:* |
|---|---|
| | mrate |
| Legal drinking age (causal effect) | 10.804** |
| | (4.479) |
| Observations | 714 |
| $R^2$ | 0.986 |
| Adjusted $R^2$ | 0.985 |
| Residual Std. Error | 17.339 (df = 649) |
| F Statistic | 726.005*** (df = 65; 649) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

$\Rightarrow$ *What did we find?*

# The parallel trends assumption

# Main assumption

**Units can be different, but – absent treatment – they must follow the same trend** (hence the panel data).

# Main assumption

**Units can be different, but – absent treatment – they must follow the same trend** (hence the panel data).

- The regression assumes a counterfactual

# Main assumption

**Units can be different, but – absent treatment – they must follow the same trend** (hence the panel data).

- The regression assumes a counterfactual $\rightarrow$ remember the extrapolation.

# Main assumption: The way around

**When we have several treated and control units they can follow .**

# Main assumption: The way around

**When we have several treated and control units they can follow** .

- individual trend lines. . .

# Main assumption: The way around

**When we have several treated and control units they can follow .**

- individual trend lines. . .
- . . . that are modelled as deviations from one unique trend

# Main assumption: The way around

**When we have several treated and control units they can follow .**

- individual trend lines. . .
- . . . that are modelled as deviations from one unique trend

# Main assumption: The way around

**When we have several treated and control units they can follow** .

- individual trend lines...
- ... that are modelled as deviations from one unique trend

⇒ *We do that with an interaction effect!*

```
mod <- lm(mrate ~ 0 +
            legal +
            state *
            year_fct,
          df)
```

###Last fix **If our units are in fact several units** (say, populations in states)

# Main assumption: The way around

**When we have several treated and control units they can follow .**

- individual trend lines...
- ... that are modelled as deviations from one unique trend

⇒ *We do that with an interaction effect!*

```
mod <- lm(mrate ~ 0 +
            legal +
            state *
            year_fct,
          df)
```

###Last fix **If our units are in fact several units** (say, populations in states)

- we can use weights

# Main assumption: The way around

**When we have several treated and control units they can follow .**

- individual trend lines…
- … that are modelled as deviations from one unique trend

⇒ *We do that with an interaction effect!*

```
mod <- lm(mrate ~ 0 +
            legal +
            state *
            year_fct,
          df)
```

###Last fix **If our units are in fact several units** (say, populations in states)

- we can use weights

⇒ *There's a tradeoff: treatment is at the unit level, statistical power at the subunit level.*