

# Econometrics Summary

---

## Slides lecture 1:

Before/after comparisons may under/overestimate treatment effects due to naive counterfactuals (pokon example).

Naive counterfactuals:

- Before/After comparison
- Cross-sectional comparison (selection bias)

Selection bias = omitted variable bias

Potential outcomes:

- Outcome for subject/group when treated
- Outcome for subject/group when not treated

Problem of causal interference: **only one potential outcome is observed**

Possible solution: perfectly comparable control group as proxy for the unobserved potential outcome.

Selection bias may be caused by self-selection

An omitted variable should:

1. differ between treatment and control groups.
2. be related to both the outcome and treatment variable.
3. be unobserved.

Individual  $i$  has two potential outcomes:

1.  $Y_{1i}$  : With treatment
2.  $Y_{0i}$  : Without treatment

The causal effect of being subjected to treatment now becomes:  $Y_{1i} - Y_{0i}$

Comparable control groups are generated through **randomized trial**.

*Check for balance*: after randomization, are differences in characteristics of control and treatment group statistically significantly different from zero? Lack of balance is an indication of failure of randomization.

---

Statistical significance  $\neq$  economic significance

Is the effect small or large?

$$100\% \cdot \hat{\beta} / \text{baseline mean of outcome variable}$$

Mechanism experiment to:

1. Single out one mechanism and see how strongly it works
2. As a screening device

## Slides lecture 2:

---

**Instrumental variables** can be used to isolate random variation in treatment.

**ITT**: Intention To Treat – The effect of *prescribing* the treatment rather than *taking* the treatment.

- *Treatment dilution*: subjects assigned to treatment fail to take treatment
- *Treatment migration*: subjects in the control group take the treatment

What we are looking for is the effect of *treatment delivered* → isolate random variation in treatment delivered with an instrumental variable.

Three assumptions when using the IV-approach:

1. Independence assumption: instrument isolates as good as random variation in treatment.
2. First stage: instrument should be sufficiently strong, the share of random variation in treatment delivered should be sufficiently large.
3. Exclusion restriction: instrument should only affect the outcome by way of the treatment

**LATE**: Local Average Treatment Effect – The effect of treatment on *compliers*, and not the effect on never/always-takers.

*Important note*: LATE is not always representative of ATE, never/always-takers may respond differently.

---

### The IV-approach:

First stage effect – effect of the instrument ( $Z$ ) on treatment dummy ( $D$ ) :  $\phi$

- $Y_i = \alpha + \beta D_i + \varepsilon_i$

Reduced form (*ITT*) – effect of the instrument on the outcome variable ( $Y$ ) :  $\rho$

- $D_i = \alpha + \beta Z_i + \varepsilon_i$

Second stage effect (*LATE*):  $\lambda = \rho \cdot \frac{1}{\phi}$

- $Y_i = \alpha + \beta \widehat{D}_i + \varepsilon_i$

As soon as  $\phi < 1$ , there is imperfect compliance and thus  $LATE > ITT$

The instrumental variable approach comes in handy when compliance is imperfect, it can be used to show treatment effect for compliers by isolating the random part of variation in treatment delivered.

### Slides lecture 3:

#### Regression discontinuity

- Treatment dummy: 0 before treatment, 1 after treatment.
- Control group and treatment group are expected to be very similar around the cutoff and only differ in terms of receiving the treatment

RD Estimation equation:

$$Y_i = \alpha + \beta \cdot cy_i + \rho \cdot D_i + \varepsilon_i$$

Where  $cy_i$  is the 'running variable' and  $D_i$  is the treatment dummy

- We expect as-good-as random variation in treatment due to assignment based on one observable variable (as long as they do not differ in other terms)

There is thus no omitted variable bias as long as covariates change continuously and smoothly  $\rightarrow$  then the only discontinuous change is caused by the treatment.

The regression discontinuity design provides a *local* average treatment effect: the effect identified is specific to the group of subjects around the cutoff. People further from the cutoff may differ from this group.

The relationship between the outcome and the running variable does not need to be flat. Dissimilarities between treatment and control may differ as long as the relationship between the running variable and the outcome can be captured.

Even situations with polynomial or changing relationships can be dealt with.

**Bandwidth:** the width of the window around the cutoff

The choice of the bandwidth is a tradeoff between noise and bias:

*Narrow (Broad) bandwidth:*

- + : Less (more) concern about choosing functional forms/non-linear trends.
- - : Estimate of the treatment effect becomes more (less) noisy.

**Important: Manipulation assumption** — Subjects cannot manipulate treatment assignment

**Falsification test**— Test for an effect of the treatment on an outcome that you expect to be **unaffected** by the treatment.

- $A_i = \alpha = \beta \cdot X_i + \rho \cdot D_i + \varepsilon_i$

Where  $A$  is that other outcome,  $X$  is the running variable and  $D$  is the treatment dummy.

---

**Fuzzy RD**: when the treatment does not sharply turn on/off at the cutoff, but *treatment intensity* changes. Or *probability of receiving treatment* changes

- $Y_i = \alpha = \beta \cdot R_i + \lambda \cdot T_{del\ i} + \varepsilon_i$

Where  $T_{del}$  stands for treatment delivered and  $R$  is the running variable.

Fuzzy RD: **Combine with IV** → isolate random variation in  $T_{del}$ , use original treatment assigned ( $D$ ) as an instrument for  $T_{del}$

- First stage:  $T_{del} = \alpha + \beta \cdot R_i + \phi \cdot D_i + \varepsilon_i$
- From first stage: predict  $\widehat{T}_{del}$
- Second stage:  $Y_i = \alpha + \beta \cdot R_i + \lambda \cdot \widehat{T}_{del\ i} + \varepsilon_i$

## Slides lecture 4

---

**Difference-in-differences**: Two groups, two periods

Control group needs to have a similar trends, levels may differ → Treatment effect: difference between treatment and control in difference between before/after treatment levels.

*DiD estimation equation:*

$$Y_{it} = \alpha + \beta \cdot (TREAT_i \cdot POST_t) + \delta \cdot TREAT_i + \gamma \cdot POST_t + \varepsilon_{it}$$

Where  $TREAT$  is equal to 1 for the treatment group and 0 otherwise. Where  $POST$  is equal to 1 for the treatment period and 0 otherwise. And where  $TREAT \cdot POST$  is the treatment dummy, equal to 1 for treatment group during treatment period, 0 otherwise.

The  $TREAT$  variable takes away level differences between the treatment and control group.

The  $TREAT \cdot POST$  variable tests whether trend post-treatment in treatment group differs from trend in control group.

Estimated treatment effect:  $\beta$

---

Framework can be easily extended to include more than two groups and periods.

*Multiple groups/periods estimation equation:*

$$Y_{it} = \alpha + \beta \cdot Policy_{it} + \delta_i + \gamma_t + \varepsilon_{it}$$

Where  $Policy_{it}$  is the treatment dummy, 1 for the treatment group during treatment period, 0 otherwise.

Where  $\delta$  represents group fixed effects, a series of dummies: one for each group.

And where  $\gamma$  represents time fixed effects, controlling for the common trend using a series of dummies: one for each time period.

Not all (external) factors may stay fixed after start of treatment: include **observed covariates** that change differently over time for each group → only works for observed variables

One possible solution: including *group-specific linear time trends*. This is not a reason to neglect the common trends assumption → only an additional check on results.

*Estimation equation including covariates and trends:*

$$Y_{it} = \alpha + \beta \cdot Policy_{it} + \delta \cdot X_{it} + \mu_i + \gamma_t + \theta_{it} + \varepsilon_{it}$$

Where  $X$  represents observed covariates that vary by  $i$  and  $t$

And where  $\theta$  stands for a series of group specific linear time trends, allowing for different trends in the outcome between the groups.

DiD can be used to estimate treatment effect of *randomized trial* as well as *observational data*.

With observational data: provide story for the plausibility of:

1. Independent treatment assignment
2. Similar outcome trends
3. No other policies pursued

## Slides lecture 5

---

How to deal with **time varying treatment effects**

1. *Ignore*: simply report ATE for period elapsed since start of treatment.
2. *Impose parametric model*
3. *Graphical exploration of time-variation in treatment effect*

Ignoring time variation:

- No idea of effects on shorter/longer term
- If short-term looks worse than long term: false negative
- If short-term looks better than long term: false positive

Imposing a parametric model:

- Linear model:  $Y_{it} = \alpha \cdot T_{it} + \alpha' \cdot T_{it} \cdot t + \beta + \varepsilon_{it}$

Where  $T_{it} \cdot t$  is the interaction term of the treatment dummy and calendar time.

- This model may be misspecified (i.e. indefinite growth / convergence)
- In case of mis-specification: risk of false positives or false negatives

Graphical exploration of time variation:

- First plot, then choose specification to test treatment effect.

Two ways of producing graphical evidence:

- Multiple treatment-control comparisons: stable control group – as good as random treatment.

Treatment dummies for every time period to estimate difference in treatment vs. control at every time interval (excluding one):

$$Y_{it} = \alpha_0 + T_i T_1 \alpha_1 + T_i T_2 \alpha_2 + \dots + T_i T_z \alpha_z + \lambda_i + \mu_t + \varepsilon_{it}$$

$$Y_{it} = \alpha_0 + \sum_{t=1}^Z T_i T_t \alpha_t + \lambda + \mu_t + \varepsilon_{it}$$

Where  $T_i$  is 1 for treatment locations and 0 otherwise.

And where  $\alpha_t$  are coefficients of interest  $\rightarrow$  plot  $\alpha_t$  by time

- Time-to-event analysis: disappearing control group – as good as random timing of treatment

Randomized timing of treatment: exposure to treatment = **common event**

**Event time** ( $\tau$ ): time relative to start of the treatment

*Estimation equation using event time dummy variables (exclude one dummy)*

$$Y_{it} = \beta_0 + W_{-10} \alpha_{-10} + W_{-9} \alpha_{-9} + \dots + W_9 \alpha_9 + W_{10} \alpha_{10} + \lambda_i + \mu_t + \varepsilon_{it}$$

$$Y_{it} = \beta_0 + \sum_{\tau=-T}^T \alpha_{\tau} W_{\tau} + \lambda_i + \mu_t + \varepsilon_{it}$$

Where event time coefficients show the level of outcome at event time  $\tau$  relative to event time  $\tau = -1$ .

Where  $\mu_t$  controls for all common shocks (time-fixed effects)

And where  $\lambda_i$  represents location fixed effects.

**Note:** number of observations per event time varies (fewer at tails)

Dealing with tails: create **bins**:

$$Y_{it} = \beta_0 + W_{-10 \text{ to } -8} \alpha_{-10 \text{ to } -8} + W_{-7} \alpha_{-7} + \dots + W_{8 \text{ to } 10} \alpha_{8 \text{ to } 10} + \lambda_i + \mu_t + \varepsilon_{it}$$

Where  $W_{8 \text{ to } 10}$  is 1 for event times  $\tau = \{8, 9, 10\}$  and 0 otherwise.

**Testing significance:**

- On ATE if it is a good approximation
- Separate tests for subperiods
- Using parametric model: appropriate functional form

## Slides lecture 6

---

**Heterogeneous treatment effects:** treatment effects that vary across subjects

In RCT with **one observation** after treatment:

$$Y_i = \alpha_0 + \alpha_1 T_i + \varepsilon_i$$

*Note: no  $t$ -subscripts*

**Multiple observations:**

$$Y_{it} = \alpha_0 + \alpha_1 T_{it} + \lambda_i + \mu_t + \varepsilon_{it}$$

*Strategies to deal with treatment heterogeneity*

1. *Ignore*: simply report ATE (weighted average across subgroups)
2. Ex-post: *treatment-by-covariate* interaction
3. Ex-ante: factorial design of RCT + treatment-by-covariate interaction

Consequences of ignoring heterogeneity:

1. False negative
2. Not enabling **targeting** of treatment
3. Less idea of the **mechanism** that is at work

Including interaction term:

- Include all possible combinations (of  $T$  and  $HIGH$ )
- Fully interacted model:

$$Y_{it} = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i \cdot HIGH_i + \alpha_3 HIGH_i + \varepsilon_i$$

Where  $HIGH_i$  is the **pre-treatment** condition for unit  $i$  that allows for a different intercept when  $HIGH = 1$

And where  $T_i \cdot HIGH_i$  allows for a different slope when  $HIGH = 1$

- Marginal effect of treatment  $T$  :  $\frac{\partial Y}{\partial T} = \alpha_1 + \alpha_2 \cdot HIGH_i$
- Conditional average treatment effect:
  - If  $HIGH = 0 \rightarrow \alpha_1$
  - If  $HIGH = 1 \rightarrow \alpha_1 + \alpha_2$
- Alternative specification:

$$Y_i = \alpha_0 + \alpha_1 T_i \cdot (1 - HIGH_i) + \alpha_4 T_i \cdot HIGH_i + \alpha_3 \cdot HIGH + \varepsilon_i$$

Where  $\alpha_1$  is the treatment effect for the *LOW* group

And where  $\alpha_4$  is the treatment effect for the *HIGH* group

Notes:

1.  $\alpha_1$  and  $\alpha_4$  can differ in terms of the standard errors as well
2. Additionally test whether  $\alpha_1$  and  $\alpha_4$  differ from each other

For Dif-in-Dif:

Recall DiD estimation equation:

$$Y_{it} = \alpha_0 + \alpha_1 T_i P_t + \alpha_2 T_i + \alpha_3 P_t + \varepsilon_{it}$$

After interacting with *HIGH* and including all terms:

$$Y_{it} = \alpha_0 + \alpha_1 T_i P_t + \alpha_2 T_i + \alpha_3 P_t + \varepsilon_{it} + \alpha_4 T_i P_t HIGH_i + \alpha_5 T_i HIGH_i + \alpha_6 P_t HIGH_i + \alpha_7 HIGH_i + \varepsilon_{it}$$



Where:	Allows for estimation of
$\alpha_0$	Intercept of the <i>LOW</i> group
$\alpha_1$	Conditional average treatment effect at <i>LOW</i> group
$\alpha_2$	Level difference between treatment and control groups for <i>LOW</i> group
$\alpha_3$	Trend in the control group for <i>LOW</i> group
$\alpha_4$	Conditional average treatment effect for <i>HIGH</i> group
$\alpha_5$	Level difference between treatment and control group for <i>HIGH</i> group
$\alpha_6$	Trend in the control group for <i>HIGH</i> group
$\alpha_7$	Intercept of the <i>HIGH</i> group

Here the conditional average treatment effect is  $\alpha_1 + \alpha_4$  for the *HIGH* group and  $\alpha_1$  for the *LOW* group.

For Dif-in-Dif: multiple groups, multiple periods

- Differential treatment effect is again estimated by:  $\alpha_4 T_i P_t \cdot HIGH_i$
- Subject fixed effects are captured by:  $\alpha_5 T_i \cdot HIGH_i + \alpha_7 \cdot HIGH_i$
- $\alpha_6 P_t \cdot HIGH_i$  allows for different trends in control groups in *HIGH* and *LOW* groups

---

**Continuous interaction variables** (rather than dichotomous)

- *HIGH* could take any value between 0 and 1
- Marginal effect of treatment also continuously varying with *HIGH*  $\rightarrow \frac{\partial Y}{\partial T} = \alpha_1 + \alpha_2 \cdot HIGH_i$

Possibility 1 Apply fully interacted model

Reporting  $\alpha_1$  and  $\alpha_2$  and thus only reporting treatment effect at *HIGH* = 0 and the difference in effect when *HIGH* = 1

Possibility 2 Applying the alternative specification

Reporting  $\alpha_1$  and  $\alpha_4$  and thus only reporting treatment effect at *HIGH* = 0 & *HIGH* = 1 and no intermediate values of *HIGH*

*Note: HIGH* = 0 & *HIGH* = 1 may be totally uninteresting

Possibility 3 Creating a marginal effect plot

Showing the marginal effect of the treatment at all actual values of the interaction variable

---

Without random assignment into the groups, there is no guarantee of balance and therefore the interaction term may not have **causal interpretation** as *LOW/HIGH* groups may differ on other (un)observable characteristics.

The interpretation may predict but not cause treatment effect.

Data mining: (ex-post)

**Bonferroni correction:** divide target  $p$ -value by number of conducted hypothesis tests.

Because of the increasing chance of finding an effect as the number of tested hypotheses increases.

Factorial design: (ex-ante).

The idea of randomizing treatment **and** interaction term to allow for causal interpretation of the interaction term.

With continuous interaction term: **matched pair randomization**

- Rank subjects from highest to lowest pre-treatment value of interaction term
- Randomly assign one subject per pair to treatment