Identification (2)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

- **Topics:**
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. Spatial econometrics
     - Spatial data, autocorrelation, spatial regressions
  3. Identification
     - Research design, IV, OLS, RDD, quasi-experiments, standard errors
  4. Hedonic pricing
     - Theory and estimation
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

VU VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

- **8 steps when undertaking research**

1. Formulate your hypotheses
2. Determine the 'treatment' variable(s) and the 'outcome' variable(s)
3. **Think of an identification strategy to identify causal effects**
4. Select samples, discuss measurement error and provide descriptives
5. Determine functional form of variables of interest
6. Think of different issues in estimating standard errors
7. Estimate model and interpret the results
8. Provide robustness checks of the results

VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

- In economics, identification of causal effects is of key importance
  - Step 3 is key → *"think of an identification strategy to identify causal effects"*

- Possible identification strategies
  1. Randomised experiments
  2. Exhaustive set of controls
  3. Instrumental variables (IV)
  4. Quasi-experiments (QE)
     ↳ Regression-discontinuity designs (RDD)

2. Randomised experiments

- In most economic studies, RCTs are not applied
  - No experimental setting possible
  - Ethical reasons / fairness
  - Costly
  - Expected substantial heterogeneity in outcomes
  - Hard to measure long-run effects
  - Lab setting may bias outcomes
    - » Recall: biases in Stated Preference surveys

2. Randomised experiments

- A more philosophical critique on RCTs
  - We might find a causal effect of x on y, but <u>do not know *why* there is an effect</u>

- Theoretical models and reasoning are needed to explain *why* we would expect a causal effect
  - Deaton (2010)

VU VRIJE UNIVERSITEIT AMSTERDAM

3. OLS with controls

- **Possible identification strategies**
  1. Randomised experiments
  2. <u>Exhaustive set of controls</u>
  3. Instrumental variables (IV)
  4. Quasi-experiments (QE)
     ↳ Regression-discontinuity designs (RDD)

VRIJE UNIVERSITEIT AMSTERDAM

3. OLS with controls

- Use an exhaustive set of controls
  - In some applications, you might know all explanatory variables

- For example, computers?
  - You aim to know the willingness to pay for a new processor
  - $price_i = \rho(processor\ quality_i) + (characteristics)_i'\gamma + \eta_i$

- Not all characteristics are available in the data
  - Houses, cars, etc.
  - Omitted variable bias…

VU VRIJE UNIVERSITEIT AMSTERDAM

3. OLS with controls

- **Use first-differencing or fixed effects to make this approach more convincing**
  - **Controls for all time-invariant factors**
  - **Requires '*within*' variation**

3. OLS with controls

- **First-differencing**
  $\Delta y_{it} = \Delta\alpha_t + \beta\Delta x_{it} + \gamma\Delta c_{it} + \Delta\epsilon_i$
  where $\Delta y_{it} = y_{it} - y_{it-1}$, etc.

- This controls for all time-invariant characteristics of i
  - Hence there should be variation in x_{it} over time

VU VRIJE UNIVERSITEIT AMSTERDAM

3. OLS with controls

- **Fixed effects**

$$y_{ig} = \bar{y}_i + \beta(x_{ig} - \bar{x}_g) + \gamma(c_{ig} - \bar{c}_g) + (\epsilon_{gt} - \bar{\epsilon}_g)$$
$$= \beta x_{ig} + \gamma c_{ig} + \mu_g + \epsilon_{ig}$$

where μ_g is a fixed effect at the level of group g

- Fixed effects vs. first-differencing
  - Identical to first-differencing when having two observations per group
  - Fixed effects is more efficient

What is true with respect to panel data?

If you have panel data, always use the panel dimension!

With first-differencing and fixed effects, the effect should be immediate and permanent

With fixed effects, there should be variation in the outcome variable within groups

You can use first-differencing over time, even if there is no variation in the treatment variable over time

False

True

- **Possible identification strategies**
  1. Randomised experiments
  2. Exhaustive set of controls
  3. <u>Instrumental variables (IV)</u>
  4. Quasi-experiments (QE)
     ↳ Regression-discontinuity designs (RDD)

VRIJE UNIVERSITEIT AMSTERDAM

4. IV

- There are two conditions for valid instruments

I. **Instrument relevance**: $\text{cov}[z_i, x_i] \neq 0$
   - μ should be statistically significant and strong
   - Rule-of-thumb: $F > 10$
   - Use Kleibergen-Paap F-statistic with multiple endogenous variables

II. **Instrument exogeneity**: $\text{cov}[z_i, \epsilon_i] = 0$
   - Instrument should not be correlated to error term
   - Instrument should only influence y via x
   - *Based on economic reasoning*

4. IV

- IV identifies local average treatment effect
  - Say that the instrument is only observed for a certain group, then IV identifies treatment effect for this group
  - Different instruments may lead to different β

- Example: gender of children and housing demand

Identification (2)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate

VRIJE UNIVERSITEIT AMSTERDAM
tinbergen institute

Identification (3)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

1. Introduction

- **Topics:**
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. Spatial econometrics
     - Spatial data, autocorrelation, spatial regressions
  3. Identification
     - Research design, IV, OLS, RDD, quasi-experiments, standard errors
  4. Hedonic pricing
     - Theory and estimation
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

VU VRIJE UNIVERSITEIT AMSTERDAM

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

- **Possible identification strategies**
  1. Randomised experiments
  2. Exhaustive set of controls
  3. Instrumental variables (IV)
  4. Quasi-experiments (QE)
     ↳ Regression-discontinuity designs (RDD)

2. Quasi-experiments

- Use exogenous shocks in the economy to identify causal effects
  - ‘Quasi’-experiments / natural experiments

- National policy changes, (arbitrary) policy rules, earthquakes, bombings
  - These shocks cannot be influenced by the individual decision makers
  - Recall: if shock is really random, selection effect is equal to zero
  - The research context indicates whether shock is indeed random

VRIJE UNIVERSITEIT AMSTERDAM

- Regression-discontinuity design (RDD)
  - Quasi-experimental method

- Assume that we have a treatment effect that is dependent on r_i:

$$x_i = \begin{cases} 1 & \text{if } r_i \geq r_0 \\ 0 & \text{if } r_i < r_0 \end{cases}$$

  - r_0 is some cutoff value

- This leads to a regression:

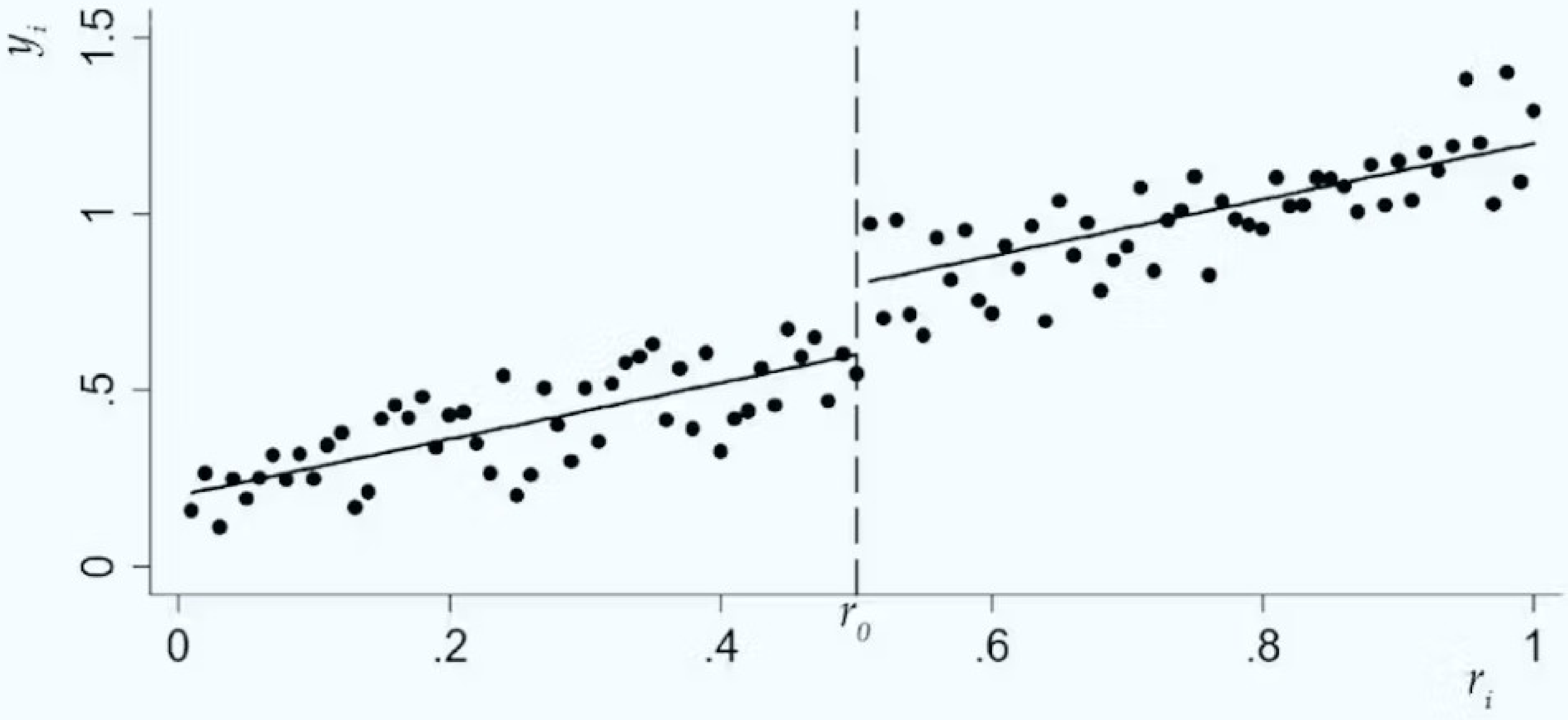
$$y_i = \alpha + \beta x_i + \gamma r_i + \epsilon_i$$

  - Note that x_i is a fully deterministic function of r_i
  - Not perfectly collinear because r_i is continuous

VU VRIJE UNIVERSITEIT AMSTERDAM

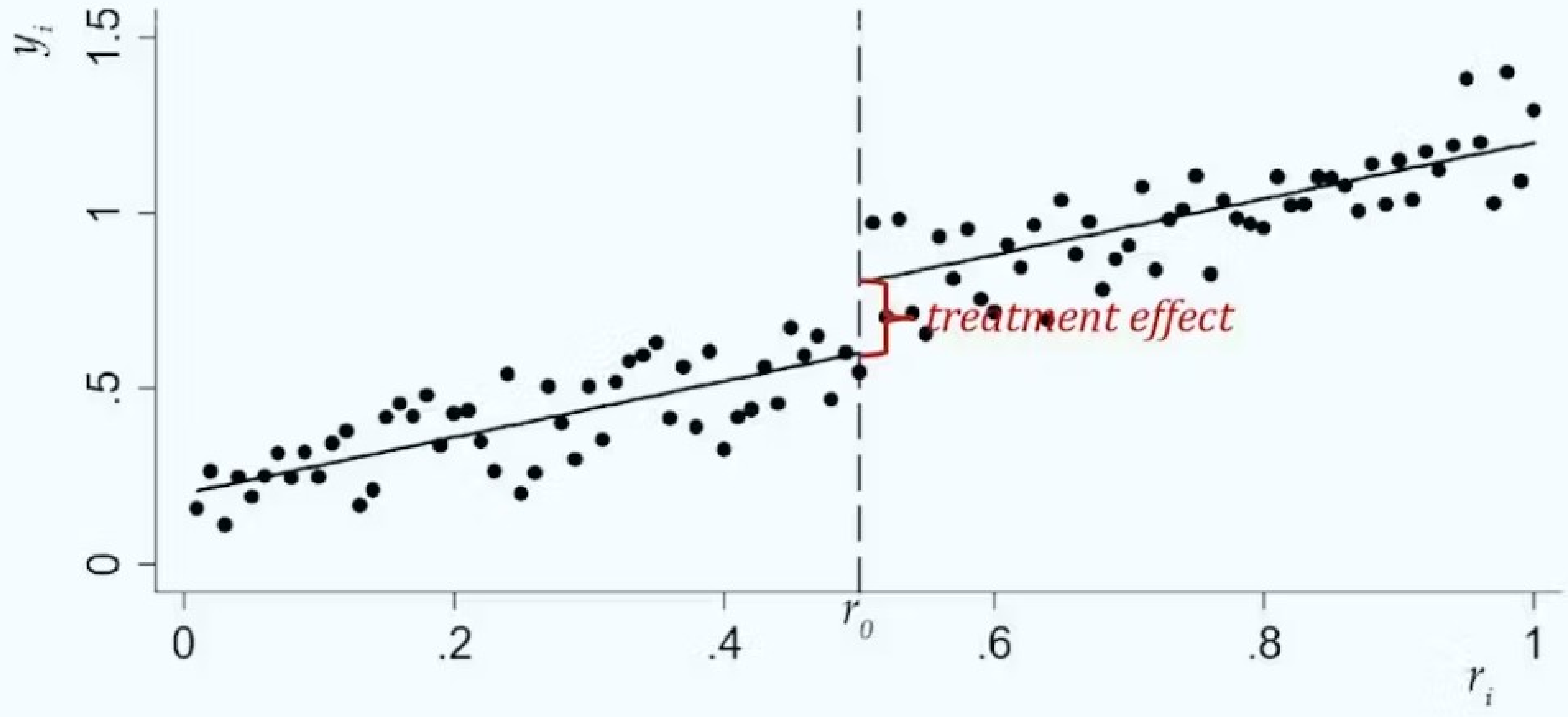
- **Example:**
  - **Students get a scholarship if they achieve a certain test-score**
  - **You aim to know the impact of the scholarship on job market outcomes**
    - » **e.g. wages**

- **Plot**



- **Control for test scores and investigate the jump in treatment at r_0**

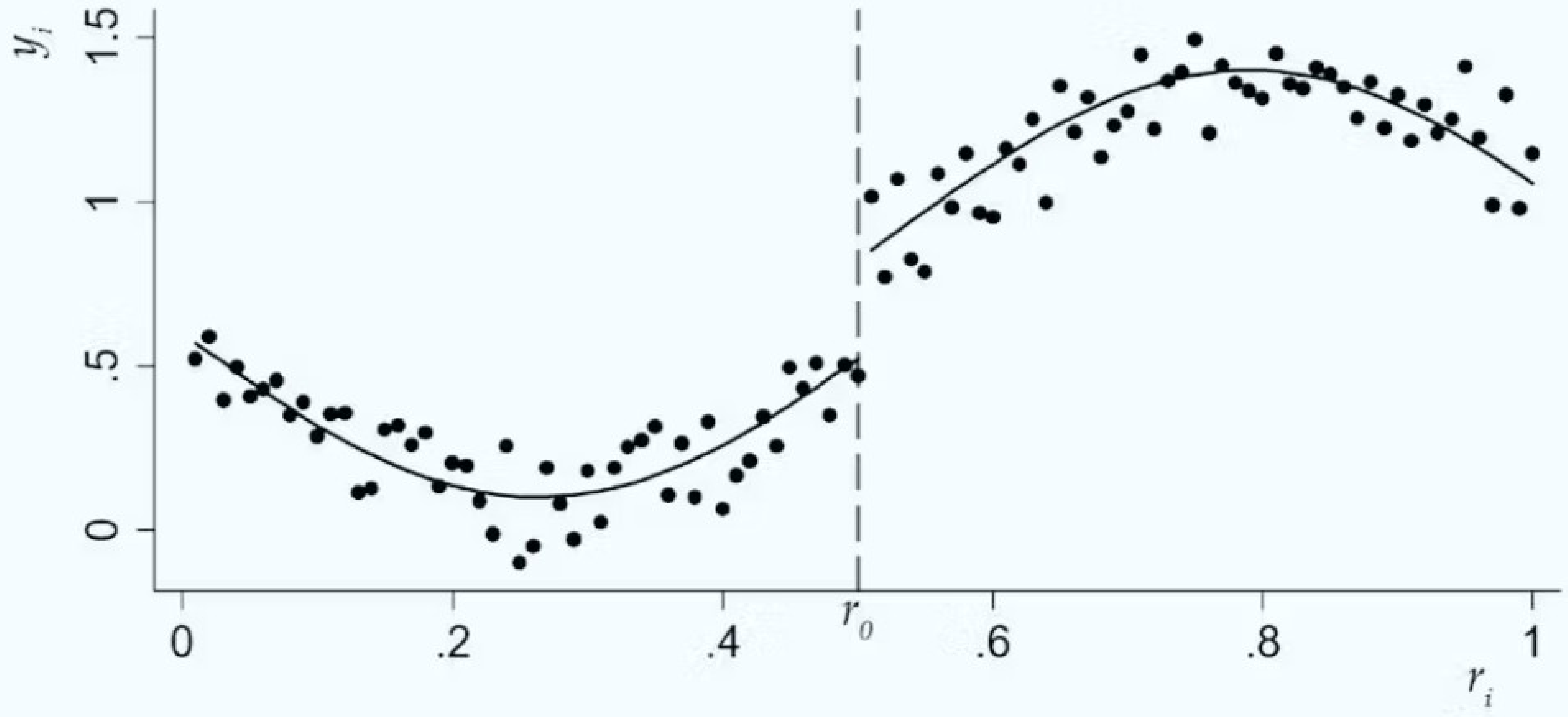
- Plot



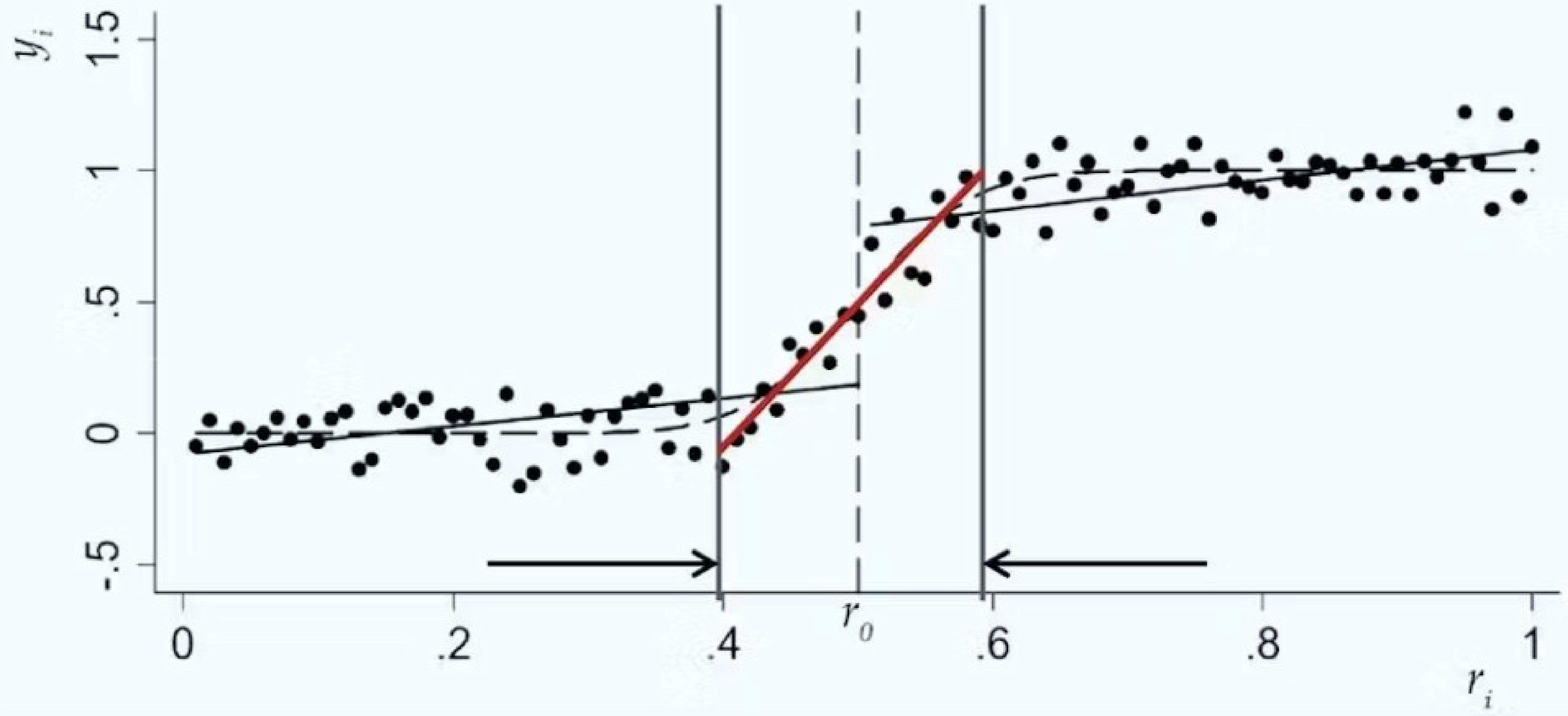
- Control for test scores and investigate the jump in treatment at r_0

VRIJE UNIVERSITEIT AMSTERDAM

- **What if x is non-linearly related to Y**
  - $y_i = \alpha + \beta x_i + f(r_i) + \epsilon_i$
  - **Specify** $f(r_i) = \gamma_1 r_i + \gamma_2 r_i^2 + \cdots + \gamma_q r_i^q$
    - q^{th}**-order polynomial**
    - **Can be estimated by OLS**



- To check for nonlinearities in a RDD is important
  - To reduce the possibility of mistakes, you may focus on observations 'close' to r_0
  - Reduces precision



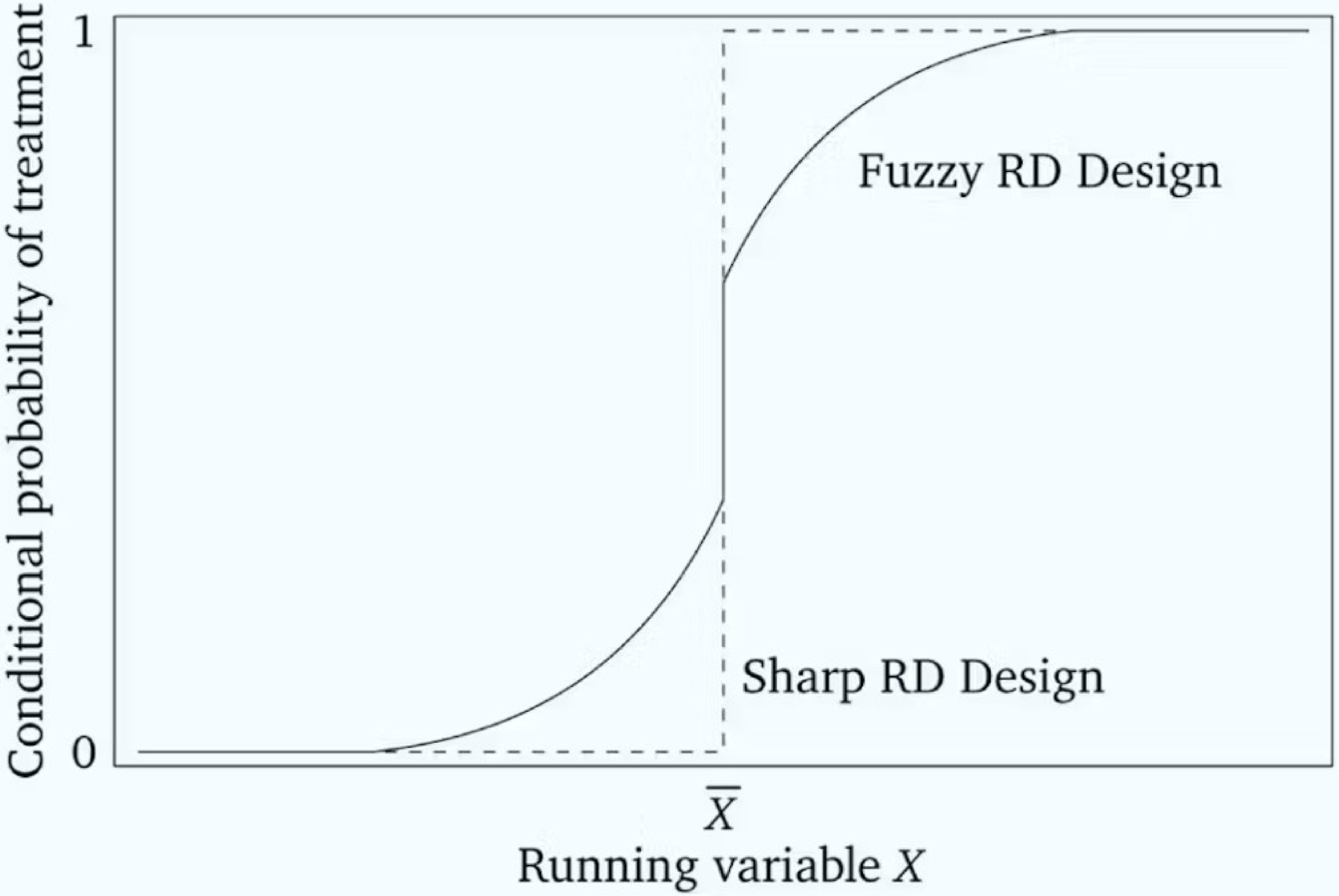
3. RDD

- Two different versions
  - *Sharp* RDD → Jump in treatment
  - *Fuzzy* RDD → Jump in probability of treatment

- Previous slides: sharp RDD

- Fuzzy RDDs are very common
  - Assignment is often 'fuzzy'

- **Illustration of a fuzzy RDD**



- **Fuzzy RDD**
  - $\text{Prob}[x_i = 1 \mid r_i] = \begin{cases} g_1(r_i) & \text{if } r_i \geq r_0 \\ g_0(r_i) & \text{if } r_i < r_0 \end{cases}$

    where $g_1(r_i) \neq g_0(r_i)$

- $\text{Prob}[x_i = 1 \mid r_i] = g_0(r_i) + [g_1(r_i) - g_0(r_i)]z_i$
  - $z_i = \mathbb{I}(r_i \geq r_0)$

- **Looks complicated – it just means that treatment probability is discontinuous at some point**

- **This leads to a two-stage least squares estimator**
  - **First stage** → $x_i = \zeta + \eta z_i + g(r_i) + \xi_i$, **with** $z_i = \mathbb{I}(r_i \geq r_0)$
  - **Second stage** → $y_i = \alpha + \beta \hat{x}_i + f(r_i) + \epsilon_i$

VU VRIJE UNIVERSITEIT AMSTERDAM

3. RDD

- Koster and Van Ommeren (2019)

- What is the impact of urban renewal programmes on house prices?
  - € 216 million by national government
  - € 1 billion by public housing associations

- Investments mainly in restructuring of public housing stock

- **Example of targeted neighbourhoods in Amsterdam:**

3. RDD

- **Use first-differencing, denoted by Δ:**

  $$\Delta y_{it} = \Delta\alpha + \beta\Delta x_{it} + \gamma\Delta c_{it} + \Delta\mu_t + \Delta\epsilon_{it}$$

  where i **property**

  t **year**

  y_{it} **log house price**

  x_{it} **in a targeted neighbourhood**

  c_{it} **control variables**

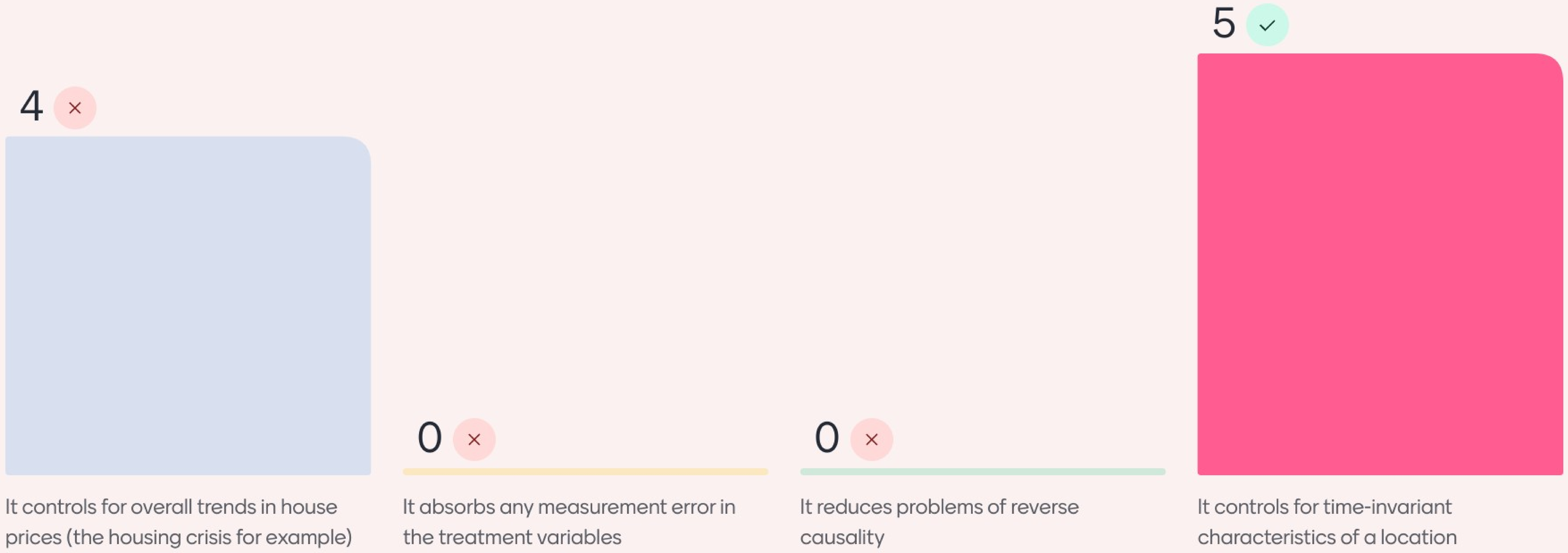
  μ_t **time fixed effects**

→ **What are the benefits of using first-differences/panel data**

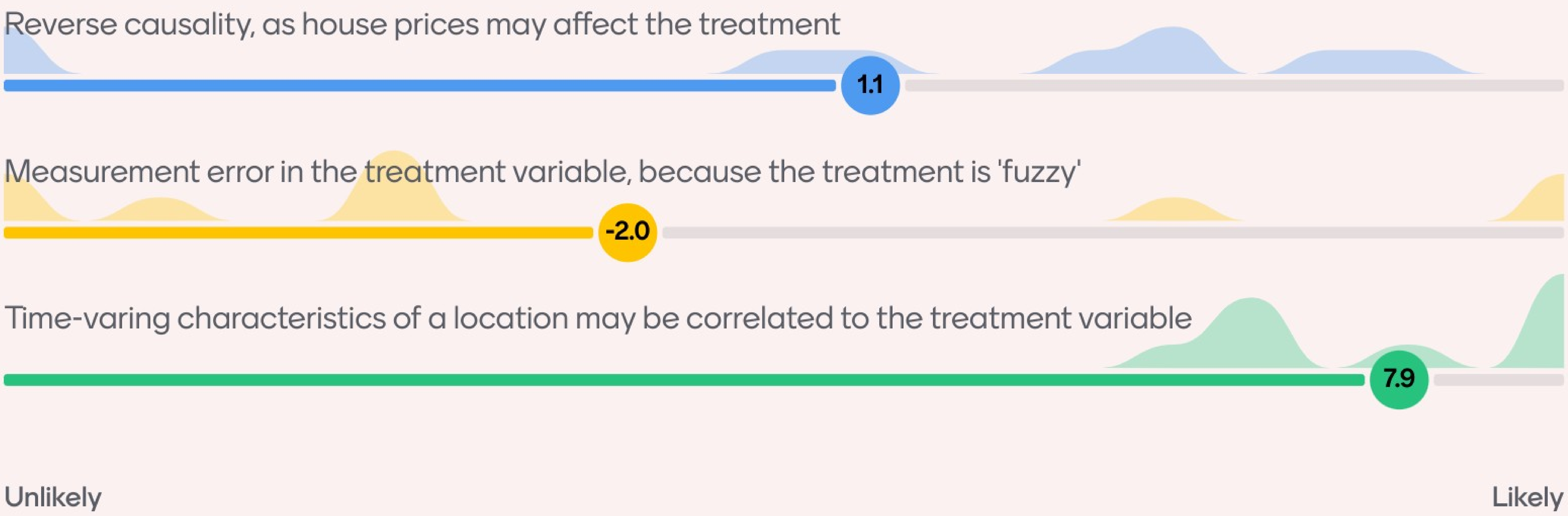
→ **What are potentially remaining endogeneity problems?**

VU VRIJE UNIVERSITEIT AMSTERDAM

What are the benefits of panel data/first-differencing in this setting?



What are potentially remaining endogeneity concerns?



9

3. RDD

- Endogeneity issue ➔ price trends of treated neighbourhoods may be different from other neighbourhoods
  - *e.g.* gentrification, trends in social interactions

- Solution: use RDD

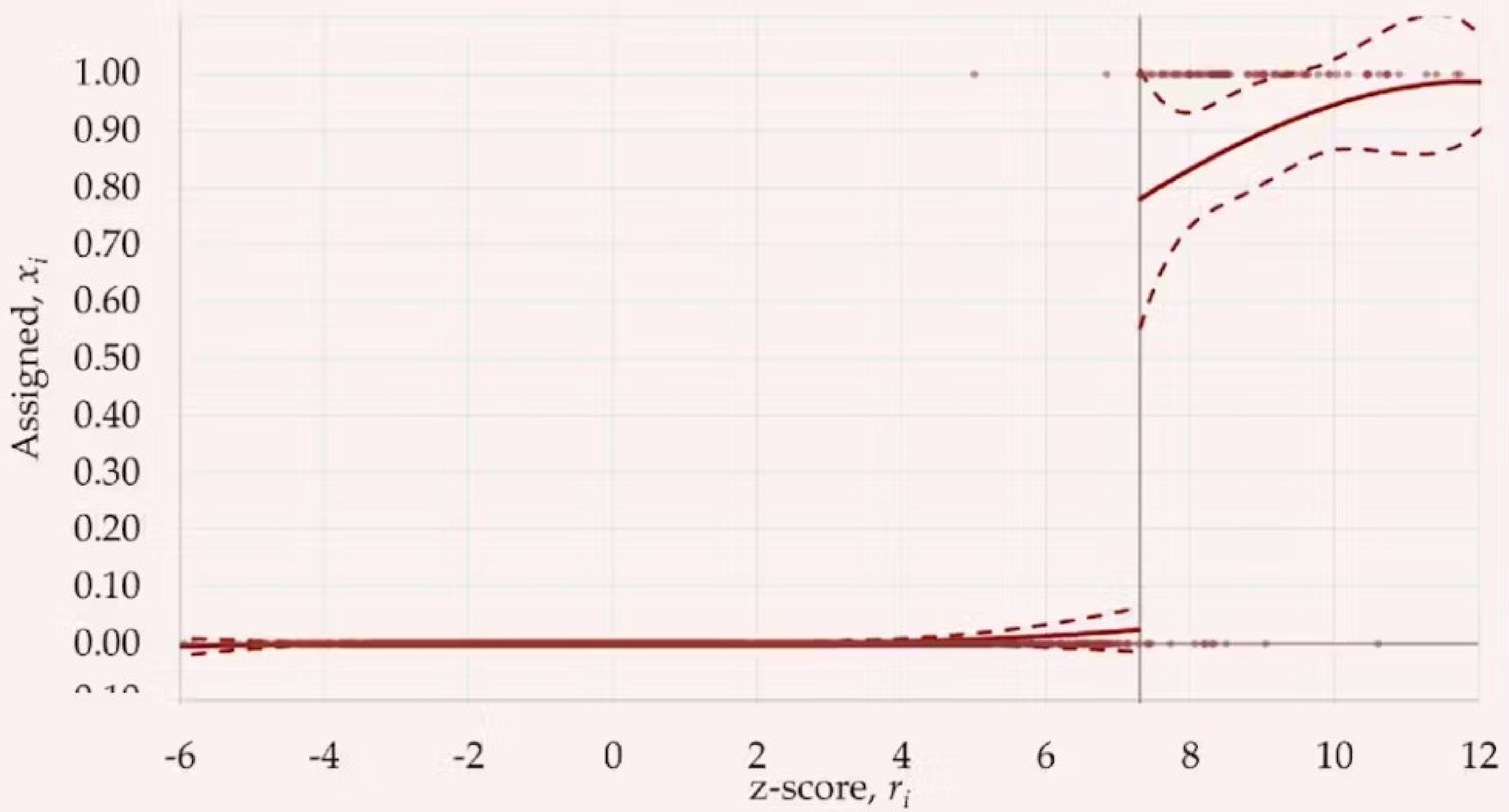
VU VRIJE UNIVERSITEIT AMSTERDAM

- **The neighbourhoods that are eligible were selected based on deprivation *z*-scores**

TABLE 1 — DEPRIVATION SCORES FOR NEIGHBOURHOODS

	All neighbourhoods		KW neighbourhoods	
	μ	σ	μ	σ
Social deprivation	0.000	0.654	1.167	0.322
Physical deprivation	0.000	0.611	2.070	0.660
Social problems	0.000	0.924	2.612	1.053
Physical problems	0.000	0.950	3.087	0.976
Overall	0.000	2.414	8.935	1.340
Number of neighbourhoods	4016		83	

- Assignment of neighbourhoods based on z-scores



→ Is this a fuzzy or a sharp RDD?

Is this a sharp or a fuzzy regression-discontinuity design?

Sharp RDD	Fuzzy RDD	This is not an RDD
0	0	0
✗	✓	✗

- **Only select observations close to the threshold** $z_{\ell t} = 7.3$

- **+'IV'-strategy**
  - $\Delta x_{it} = \zeta + \eta \Delta z_{it} + \theta \Delta c_{it} + \Delta \nu_t + \Delta \xi_{\mathbf{it}}$ *1*[st] *stage*

    where $z_{it} = \mathbb{I}(r_i \geq r_0)$

    $z_{\ell t} = 0$ **before the programme started**
  - **Use fitted value of** Δx_{it} **in second stage**

VU VRIJE UNIVERSITEIT AMSTERDAM

3. RDD

▪ Results

Table 4.4 – URBAN RENEWAL AND HOUSE PRICES
(Dependent variable: the change in the log of house prices)

	First-differences	+Fuzzy RDD
	(1)	(2)
ΔKW investment	0.0441***	0.0329***
	(0.0114)	(0.0122)
Number of observations	169,664	22,589
R^2-within	0.375	
Kleibergen-Paap F-statistic		5444
Bandwidth, δ		3.383

Notes: We exclude observations within 2.5km of targeted neighbourhoods to avoid picking up spillover effects beyond the neighbourhood boundaries. In column (3) the change in the KW investment is instrumented with the change in the eligibility based on the scoring rule. Standard errors are clustered at the neighbourhood level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

→ Please interpret the coefficients

VRIJE UNIVERSITEIT AMSTERDAM

What statements regarding the results are true?

9

3. RDD

- Urban renewal programmes have led to changes in prices of about 3-3.5%
  - Neighbourhoods have become more attractive
  - In that respect, the programme was effective
    - Total house price increase is higher than the investments

VU VRIJE UNIVERSITEIT AMSTERDAM

5. Summary

Today

- Setting up a research project
- Alternatives to RCTs
  - OLS with controls
  - IV
  - Quasi-experimental methods

5. Summary

- 8 steps when undertaking research

1. Formulate your hypotheses
2. Determine the ‘treatment’ variable(s) and the ‘outcome’ variable(s)
3. Think of an identification strategy to identify causal effects
4. Select samples, discuss measurement error and provide descriptives
5. Determine functional form of variables of interest
6. Think of different issues in estimating standard errors
7. Estimate model and interpret the results
8. Provide robustness checks of the results

VU VRIJE UNIVERSITEIT AMSTERDAM

Identification (3)

Applied Econometrics for Spatial Economics

Hans Koster
Professor of Urban Economics and Real Estate