# Using quasi-experimental methods to evaluate public policies

Hans Henrik Sievertsen
(The Danish National Centre for Social Research - SFI)
e-mail: hhs@sfi.dk

February 22, 2017

# Objective of the talk

## Objectives

- Explain the **intuition** behind the four popular methods to identify causal effects using quasi-experimental research designs.
- Explain important **limitations** and **assumptions** behind these methods.
- Present three **examples** that use these four methods.
- Present some common economist jargon.

I use green boxes to present methods

I use blue boxes to present examples

I use gray boxes to summarize

# Outcomes of the talk

## Methods: The Furious Four

1. **Regression analysis**
2. **Difference-in-differences (DiD)**
3. **Regression Discontinuity (RD)**
4. **Instrumental Variables (IV)**

## Jargon: Econ Slang

**Omitted Variable Bias (OVB)**
**Identification Strategy**
**Fixed effects**
**Compliers**
**Causal effect**
**Potential outcomes**
**Local Average Treatment Effects**

# Point of departure: Defining treatment and outcomes

**The policy question:** *Is policy x a good policy?*

## 1 What is "policy x"?

Define the **treatment**.

## 2 What is a "good policy"?

Define the **outcomes of interest**.

## **Example** Danish School Reform 2014

- **Several treatments**: longer school day, new curriculum, physical activity in classroom
- **Several objectives/outcomes**: challenge all pupils, reduce the socio-economic gradient in academic achievement, increase trust and well-being to and in the school.

# Point of departure: Defining **specific** treatment and outcomes

### **1** What is "policy x"?

Define the **treatment**.

- Treatment: Longer school days

### **2** What is a "good policy"?

Define the **outcomes of interest**.

- Outcome: Academic achievement.

# Point of departure: Defining **specific** treatment and outcomes

### 1 What is "policy x"?

Define the **treatment**.

- Treatment: Longer school days

### 2 What is a "good policy"?

Define the **outcomes of interest**.

- Outcome: Academic achievement.

### **The evaluation problem** - You can only take one road

**Alice** goes to public school, is **treated** by the reform, and has longer school days.

- We are interested in the difference in **potential outcomes** for Alice with and without longer school days.
- The difference in potential outcomes is what we define as the **causal effect** of longer school days.

**Problem:** Alice is either treated or not. She cannot be both. We never know both potential outcomes.

How do find out whether Alice would learn more without the longer school days?

We need an **identification strategy**.

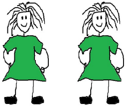
**How do find out whether Alice would learn more without the longer school days?**

## **Option 1** Clone Alice

- Give the original Alice longer school days.
- Give the cloned Alice normal school days.

⇒ Compare their learning outcomes.

**I don't want to do this.**



# Point of departure: The evaluation problem

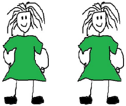
**How do find out whether Alice would learn more without the longer school days?**

## **Option 1** Clone Alice

- Give the original Alice longer school days.
- Give the cloned Alice normal school days.

⇒ Compare their learning outcomes.

**I don't want to do this.**

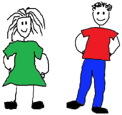


## **Option 2** Compare Alice to Bob

**Bob** goes to private school, is **not treated** by the reform, and has normal school days.

⇒ Compare Alice's learning outcome to Bob's learning outcome.

**What does this comparison tell us?**



**What does a comparison between Bob and Alice tell us?**

## What we actually observe

| | Alice | Bob |
|---|---|---|
| **Treated** | Yes | No |
| **Observed test score** | 5 | 6 |

**Observed treatment effect: 5-6=-1**

**What does a comparison between Bob and Alice tell us?**

## What we actually observe

| | Alice | Bob |
|---|---|---|
| **Treated** | Yes | No |
| **Observed test score** | 5 | 6 |

**Observed treatment effect: 5-6=-1**

## What if we observed everything?

| | Alice | Bob |
|---|---|---|
| **Treated** | Yes | No |
| Initial test score level | 3.5 | 3.5 |
| Effect of treatment | 1.5 | |
| Private tutoring | No | Yes |
| Effect of tutoring | | 2.5 |
| **Observed test score** | 5 | 6 |

**Actual treatment effect=1.5**

## Will more observations help us?

- Increasing sample size **helps in cases** with random noise.
- Increasing sample size is **not the solution** to systematic bias.

  (ex. if private schools is related to private tutoring)

**Omitted variable bias**

Omitted variable bias (OVB): we fail to control for variables that are correlated with both the treatment and outcome of interest.
$\Rightarrow$**Causes biased estimates!**

Josh Angrist: *"I get my students to rant the formula for omitted variable bias!"*

# Point of departure: The evaluation problem

## Will more observations help us?

- Increasing sample size **helps in cases** with random noise.
- Increasing sample size is **not the solution** to systematic bias.

  (ex. if private schools is related to private tutoring)

**Omitted variable bias**

Omitted variable bias (OVB): we fail to control for variables that are correlated with both the treatment and outcome of interest.
$\Rightarrow$**Causes biased estimates!**

Josh Angrist: *"I get my students to rant the formula for omitted variable bias!"*

## Solutions

1. Randomize treatment status (**Randomized Control Trials**): Actively breaks the link between treatment and omitted variables.
2. Exploit **quasi-experimental randomization**: Policies and institutional setting that causes randomization of treatment status.

# Outcomes of the talk

## Methods: The Furious Four

1. **Regression analysis**
2. **Difference-in-differences (DiD)**
3. **Regression Discontinuity (RD)**
4. **Instrumental Variables (IV)**

## Jargon: Econ Slang

**Omitted Variable Bias (OVB)** ✓
**Identification Strategy**✓
**Fixed effects**
**Compliers**
**Causal effect** ✓
**Potential outcomes** ✓
**Local Average Treatment Effects**

## We would like to know how length of the school day affect children

- Look for quasi-experimental variation in length of school day.
- I observed variation in time of the day children are tested.

**Can we use the variation in time of the day children are tested?**

First step: understand what causes the variation

! Test time depends on computer availability.

⇒ For children at the **same school**: variation "as good as random".

⇒ For children at the **different schools**: variation not random.

⇒ We have to control for school **fixed effects.**

# Quasi-experimental methods: **1** Regression Analysis

## We would like to know how length of the school day affect children

- Look for quasi-experimental variation in length of school day.
- I observed variation in time of the day children are tested.

**Can we use the variation in time of the day children are tested?**

First step: understand what causes the variation

! Test time depends on computer availability.

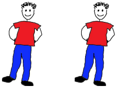
⇒ For children at the **same school**: variation "as good as random".

⇒ For children at the **different schools**: variation not random.

⇒ We have to control for school **fixed effects.**

- **Solution**: Compare children within same school.

**School 1**



| | | |
|---|---|---|
| Test time | 9 | 10 |
| Test result | 51 | 50 |
| Difference | -1 | |

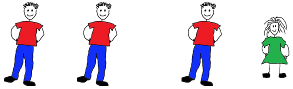
## We would like to know how length of the school day affect children

- Look for quasi-experimental variation in length of school day.
- I observed variation in time of the day children are tested.

**Can we use the variation in time of the day children are tested?**

First step: understand what causes the variation

! Test time depends on computer availability.
⇒ For children at the **same school**: variation "as good as random".
⇒ For children at the **different schools**: variation not random.
⇒ We have to control for school **fixed effects.**



| | **School 1** | | **School 2** | |
|---|---|---|---|---|
| Test time | 9 | 10 | 10 | 11 |
| Test result | 51 | 50 | 64 | 63 |
| Difference | -1 | | -1 | |

**One hour later ⇒ one point lower score.**

## Method: Regression Analysis

- Compare test results for children in **same school** with varying test times.

  **Challenge:** 1,200 schools gives approximately 1,200 results.

  **Solution:** Weighted average of results across schools ⇒ One result.

  **In practice:** Use regression analysis to compute average effect of test time on test result across schools.

⇒ **Regression analysis:** A method to hold observable factors constant.

## Example: Cognitive fatigue influences students' performance on standardized tests


PNAS

Cognitive fatigue influences students' performance on standardized tests

Hans Henrik Sievertsen[a], Francesca Gino[b,1], and Marco Piovesan[c]

[a]The Danish National Centre for Social Research, 1052 Copenhagen, Denmark; [b]Harvard Business School, Harvard University, Boston, MA 02163; and [c]Department of Economics, University of Copenhagen, 1353 Copenhagen, Denmark

Edited by Pamela Davis-Kean, University of Michigan, Ann Arbor, MI, and accepted by the Editorial Board January 15, 2016 (received for review August 25, 2015)

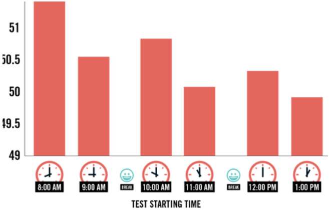
Using test data for all children attending Danish public schools between school years 2009/10 and 2012/13, we examine how the time of the test affects performance. Test time is determined by the weekly class schedule and computer availability at the school. We find that, for every hour later in the day, test performance decreases by 0.9% of an SD (95% CI, 0.7–1.0%). However, a 20- to 30-minute break improves average test performance by 1.7% of an SD (95% CI, 1.2–2.2%). These findings have two important policy implications: First, cognitive fatigue should be taken into consideration when deciding on the length of the school day and the frequency and duration of breaks throughout the day. Second, school accountability systems should control for the influence of external factors on test scores.

time of day | breaks | standardized tests | education

test. We also suggest, and find, that breaks allow students to recharge their mental resources, with benefits for their test scores.

We base these predictions on psychological research on cognitive fatigue, an increasingly common human condition that results from sustained cognitive engagement that taxes people's mental resources (4). Persistent cognitive fatigue has been shown to lead to burnout at work, lower motivation, increased distractibility, and poor information processing (5–12). In addition, cognitive fatigue is detrimental to individuals' judgments and decisions, even those of experts. For instance, in the context of repeated judicial judgments, judges are more likely to deny a prisoner's request and accept the status quo outcome as they advance through the sequence of cases without breaks on a given day (13). Evidence for the same type of decision fatigue has been found in other contexts, including consumers making choices among various alternatives (14) and physicians prescribing unnecessary antibiotics (15). Across these contexts, the overall demand of multiple decisions people face throughout the day on their cognitive resources erodes their ability to resist making potentially inappropriate or bad decisions.

ECONOMIC SCIENCES



# Outcomes of the talk

## Methods: The Furious Four

1. **Regression analysis** A method to hold observable factors constant. ✓
2. **Difference-in-differences (DiD)**
3. **Regression Discontinuity (RD)**
4. **Instrumental Variables (IV)**

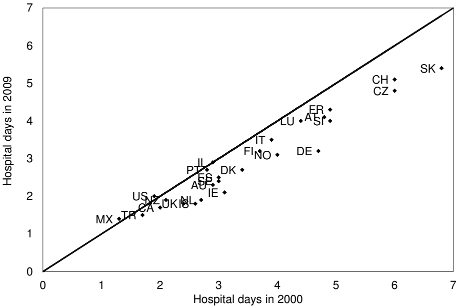
## Jargon: Econ Slang

**Omitted Variable Bias (OVB)** ✓
**Identification Strategy**✓
**Fixed effects**✓
**Compliers**
**Causal effect** ✓
**Potential outcomes** ✓
**Local Average Treatment Effects**

## Example: Care around birth. Does it matter?

- We observe variation in spending on child birth across countries and a trend towards lower spending.

**Fig:** Hospital days at birth




- Can we analyze whether reduction in care around birth has effect on child (and mother) outcomes?

## Example: Care around birth. Does it matter?

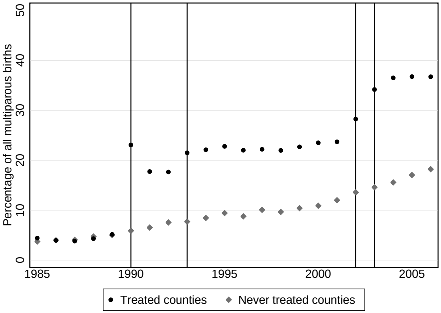
- Some stay longer at hospital and some leave immediately.
- Most variation driven by **unobservable preferences and resources**.

$\Rightarrow$ We cannot control for it.

Search for policy-driven variation in care around birth.

- **Solution:** Counties introduced mandated discharge at the day of birth.
- **Advantage:** Not all counties introduced it at the same time.

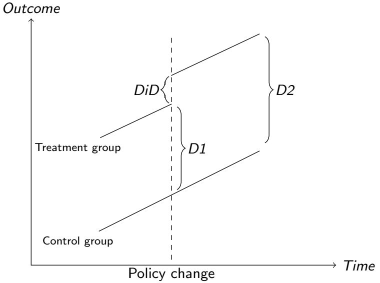
**Fig:** Hospital days at birth



## Method: Difference-in-differences (DiD)

**Intuition**

- **1** Intuition: County A introduced policy in 1990. County B introduced policy in 1993.
- **2** Mothers in County A are different than mothers in county B.
- **3** We **assume** that this difference is constant over time.
- **D1** Difference between a child in county A and a child in county B in **1989**.
- **D2** Difference between a child in county A and a child in county B in **1990**.
- **DiD** **Difference-in-differences**=D2-D1.



## Example: Care around birth. Does it matter?

**What we find**

- Treated children have short-run higher readmission rates.
- Treated children are breastfed less.
- Treated children have worse subjective health at age 7.
- Treated children have lower 9th grade GPA.
- Results are driven by "at-risk-children"

# Outcomes of the talk

## Methods: The Furious Four

1. **Regression analysis** A method to hold observable factors constant. ✓
2. **Difference-in-differences (DiD)** Compare differences between treated and untreated units, before and after the policy is introduced. ✓
3. **Regression Discontinuity (RD)**
4. **Instrumental Variables (IV)**

## Jargon: Econ Slang

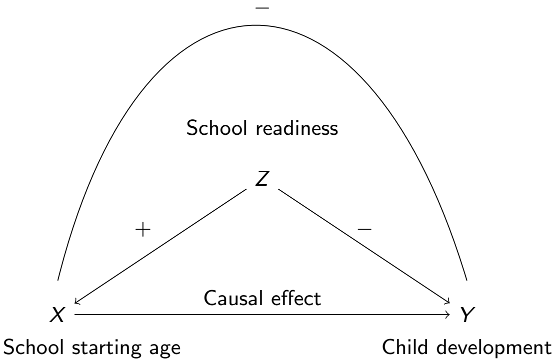
**Omitted Variable Bias (OVB)** ✓
**Identification Strategy**✓
**Fixed effects**✓
**Compliers**
**Causal effect** ✓
**Potential outcomes** ✓
**Local Average Treatment Effects**

## Example: When should children start school?

**Treatment:** Children are one year older when they start school.

**Outcome:** Children's non-cognitive skills

- Is it random who delays school? $\Rightarrow$ **No**



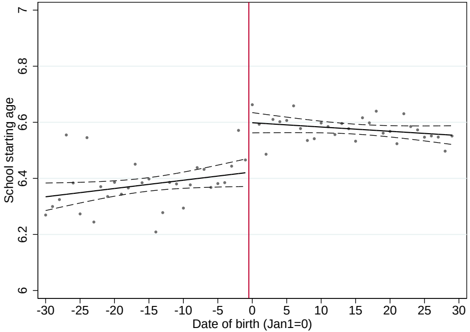
## Example: When should children start school?

- Search for policy-driven variation in school starting age (SSA).
- **Solution:** Cutoff-date: Children in DK start school the year they turn 6.

  Born Dec 31: SSA=5.6. Not treated.

  Born Jan 1: SSA=6.6. Treated.

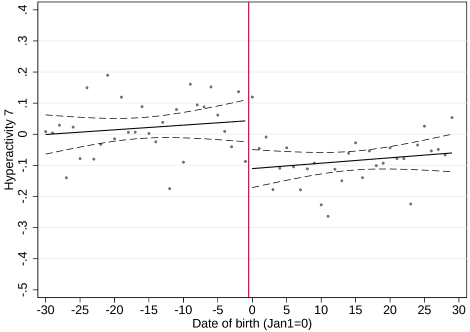
**Fig:** Date of birth and SSA



## Method: Regression Discontinuity (RD)

- A continuous curve is a smooth unbroken curve.
- Exploit the **discontinuity** in school starting age at the cutoff date.
- **Intuition:** Random (assumption) whether you are born on January 1 or December 31 and institutional setting assigns different treatment.
- The date of birth is the **forcing variable**.

**Fig:** Date of birth and Hyperactivity



# Outcomes of the talk

## Methods: The Furious Four

1. **Regression analysis** A method to hold observable factors constant. ✓
2. **Difference-in-differences (DiD)** Compare differences between treated and untreated units, before and after the policy is introduced. ✓
3. **Regression Discontinuity (RD)** Exploit that assignment to treatment changes discontinuously along a forcing variable. ✓
4. **Instrumental Variables (IV)**

## Jargon: Econ Slang

**Omitted Variable Bias (OVB)** ✓
**Identification Strategy**✓
**Fixed effects**✓
**Compliers**
**Causal effect** ✓
**Potential outcomes** ✓
**Local Average Treatment Effects**

## Example: When should children start school I (continued)?

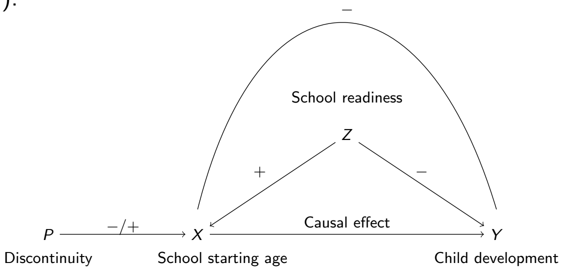
**Treatment:** Children are one year older when they start school.

**Outcome:** Children's non-cognitive skills.

**Strategy:** Exploit discontinuity in school starting age by birth date.

But not everyone complies to the rule ⇒ use discontinuity as an **instrumental variable**.

- Sharp regression discontinuity: everyone complies (no need for IV).
- Fuzzy regression discontinuity: not everyone complies (need for IV strategy).



## Method: Instrumental Variables (IV)

- **Intuition:** Find a variable that affects treatment (for example the discontinuity) $\Rightarrow$ that is our instrument.

  The instrument can only affect our outcome through the treatment variable (assumption!).
- **Step 1** Use instrument to predict variation in treatment.
- **Step 2** Estimate the relationship between the outcome of interest and the predicted variation from step 1.

**Important:** IV only identifies effect of treatment for those who change treatment status because of the instrument. The **compliers**.

## Compliers: Why is that important?

- **We are different** So far: The policy has the same effect on everyone.

  In practice, the effect of the policy may vary across people.

  **Common approach:** Focus on average treatment effect (ATE).

  **Instrumental variables** identifies the **Local Average Treatment Effect (LATE)**. Only the average effect for those who reacted to the instrument.

## Example: Optional 10th grade

- **Used IV strategy to identify effect of 10th grade**

  Instrument only affected behavior of those on in doubt!

  $\Rightarrow$ Study is **not** informative of the effect for those who were sure that they would go.

- News story: 10th grade not worth it!

## Example: When should children start school I (continued)?

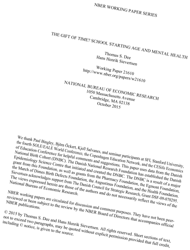
- **Intuition:** We use the discontinuous jump at January 1 as an instrument for school starting age.

  We identify the Average Treatment Effect for the compliers.

  $\Rightarrow$ 1y older SSA $\Rightarrow$ 0.7SD lower hyperactivity at age 7.

  But: LATE effect!

- Our results are not informative about:
  - children who have problems and would have delayed school enrollment if they were born in December.
  - children with high school readiness who would advanced enrollment if they were born in January.


NBER WORKING PAPER SERIES

THE GIFT OF TIME? SCHOOL STARTING AGE AND MENTAL HEALTH

Thomas S. Dee
Hans Henrik Sievertsen

Working Paper 21610
http://www.nber.org/papers/w21610

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2015

We thank Paul Bingley, Björn Öckert, Kjell Salvanes, and seminar participants at SFI, Stanford University, the fourth SOLE/EALE World Conference, the Copenhagen Education Network, and the CESifo Economics of Education Conference for helpful comments and suggestions. This paper uses data from the Danish National Birth Cohort (DNBC). The Danish National Research Foundation has established the Danish Epidemiology Science Centre that initiated and created the DNBC. The DNBC is a result of a major grant from this Foundation, as well as grants from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation. Sievertsen acknowledges support from The Danish Council for Strategic Research, Grant DSF-09-070295. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Thomas S. Dee and Hans Henrik Sievertsen. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

# Outcomes of the talk

## Methods: The Furious Four

1. **Regression analysis** A method to hold observable factors constant. ✓
2. **Difference-in-differences (DiD)** Compare differences between treated and untreated units, before and after the policy is introduced. ✓
3. **Regression Discontinuity (RD)** Exploit that assignment to treatment changes discontinuously along a forcing variable. ✓
4. **Instrumental Variables (IV)** Exploit a third variable that affects treatment status but not our outcome directly to obtain variation in treatment status. ✓

## Jargon: Econ Slang

**Omitted Variable Bias (OVB)** ✓
**Identification Strategy**✓
**Fixed effects**✓
**Compliers** ✓
**Causal effect** ✓
**Potential outcomes** ✓
**Local Average Treatment Effects** ✓

# Quasi-experimental approaches: Advantages and disadvantages

## Advantages

Often more feasible than RCT.

Often very policy relevant, because the policies are changed.

## Disadvantages

Only answer questions with quasi-experimental variation.

(Questions we can answer vs. questions we would like to answer).

Often complex assumptions.

Often very specific treatments.

Not always the relevant treatment effect.

# Where to go next?

- Contact me `hhs@sfi.dk`
- Read Mastering Metrics.