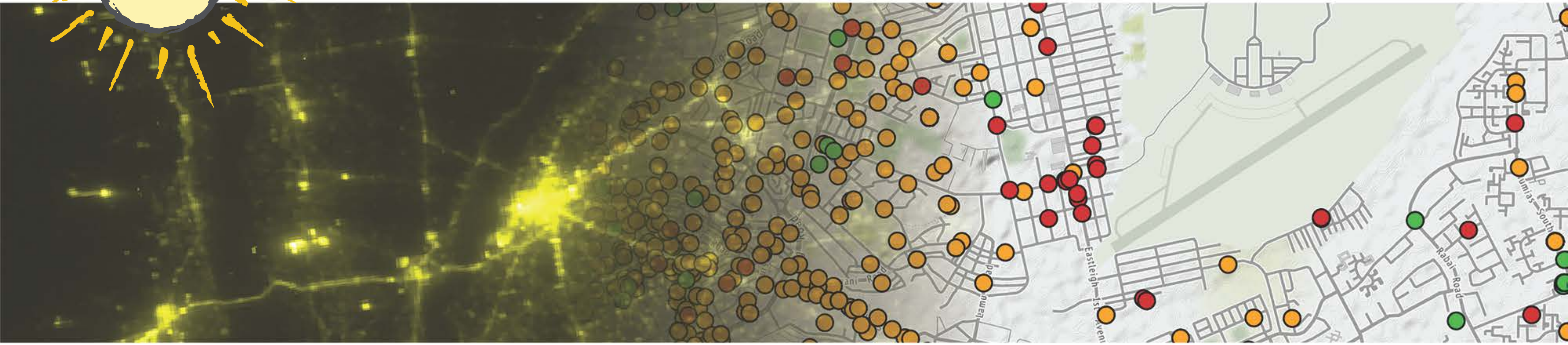
IE CONNECT FOR IMPACT

Transforming the Growth Potential of Transport Investments

Measuring Impact I: Causal Inference & Quasi-Experimental Methods

Arndt Reichert, DIME



Evaluation objective

Identify

cause - effect

enable good policy decisions

WHY IS UNDERSTANDING CAUSALITY SO IMPORTANT?

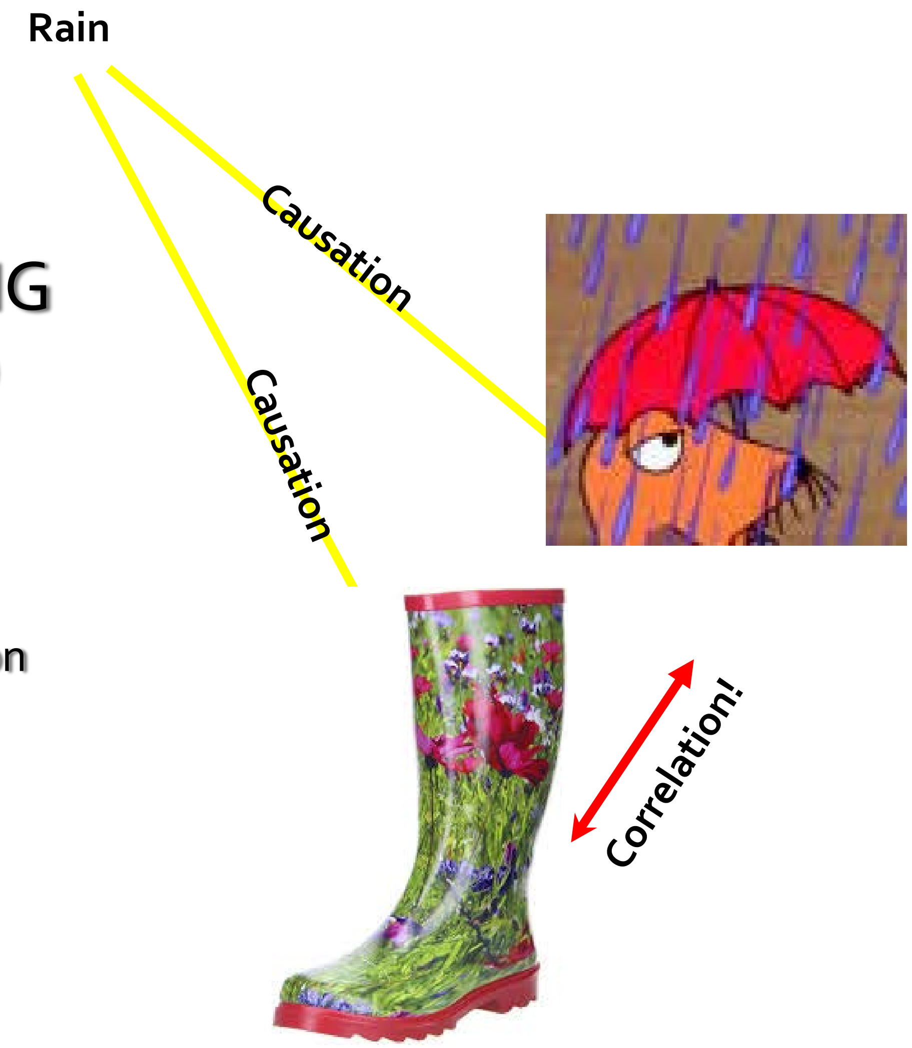
It is easy to confuse correlation with causation

WHEN I USE MY UMBRELLA MY FEET GET WET

I THINK I WILL STOP USING UMBRELLA

WHY IS UNDERSTANDING CAUSALITY SO IMPORTANT?

It is easy to confuse correlation with causation



Evaluation

Narrow down causality by identifying a counterfactual and compare

WHAT HAPPENED

TO

WHAT WOULD HAVE HAPPENED

What is counterfactual analysis?

- Compare same individual
  - with & without intervention
  - at the same point in time

Missing data

- Compare statistically identical groups of individuals
  - with & without intervention
  - at the same point in time

Comparable data

Counterfactual criteria

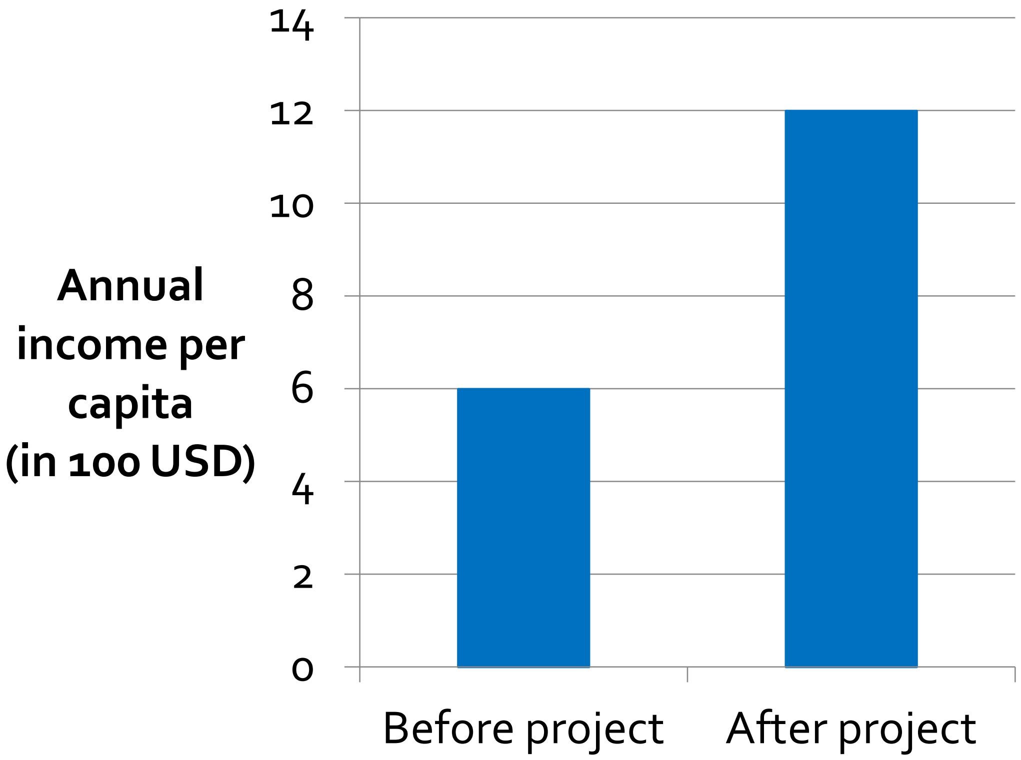
Treated & control groups

- Have identical initial average characteristics (observed and unobserved)
- So that, the only difference is the treatment
- Therefore the only reason for the difference in outcomes is due to the treatment

Counterfactual analysis

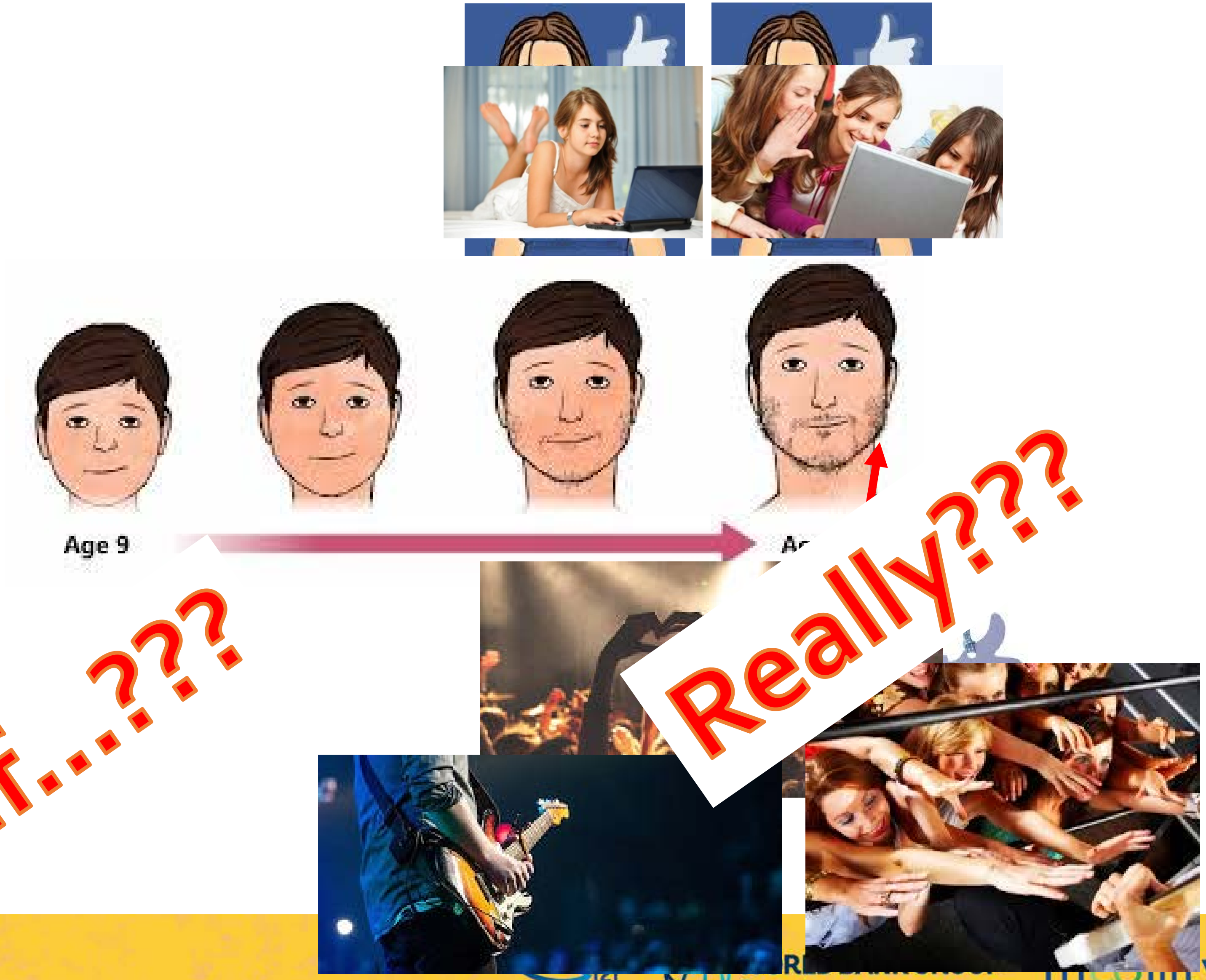
- There are methods to generate this counterfactual:
  - difference in-differences (+ matching)
  - instrumental variables
  - discontinuity design
- Each of these rely on key assumptions
- All look for a situation where potential beneficiaries were quasi randomly assigned to treatment and comparison state
- With large sample, all characteristics average out

Is Investment in Monitoring Enough?



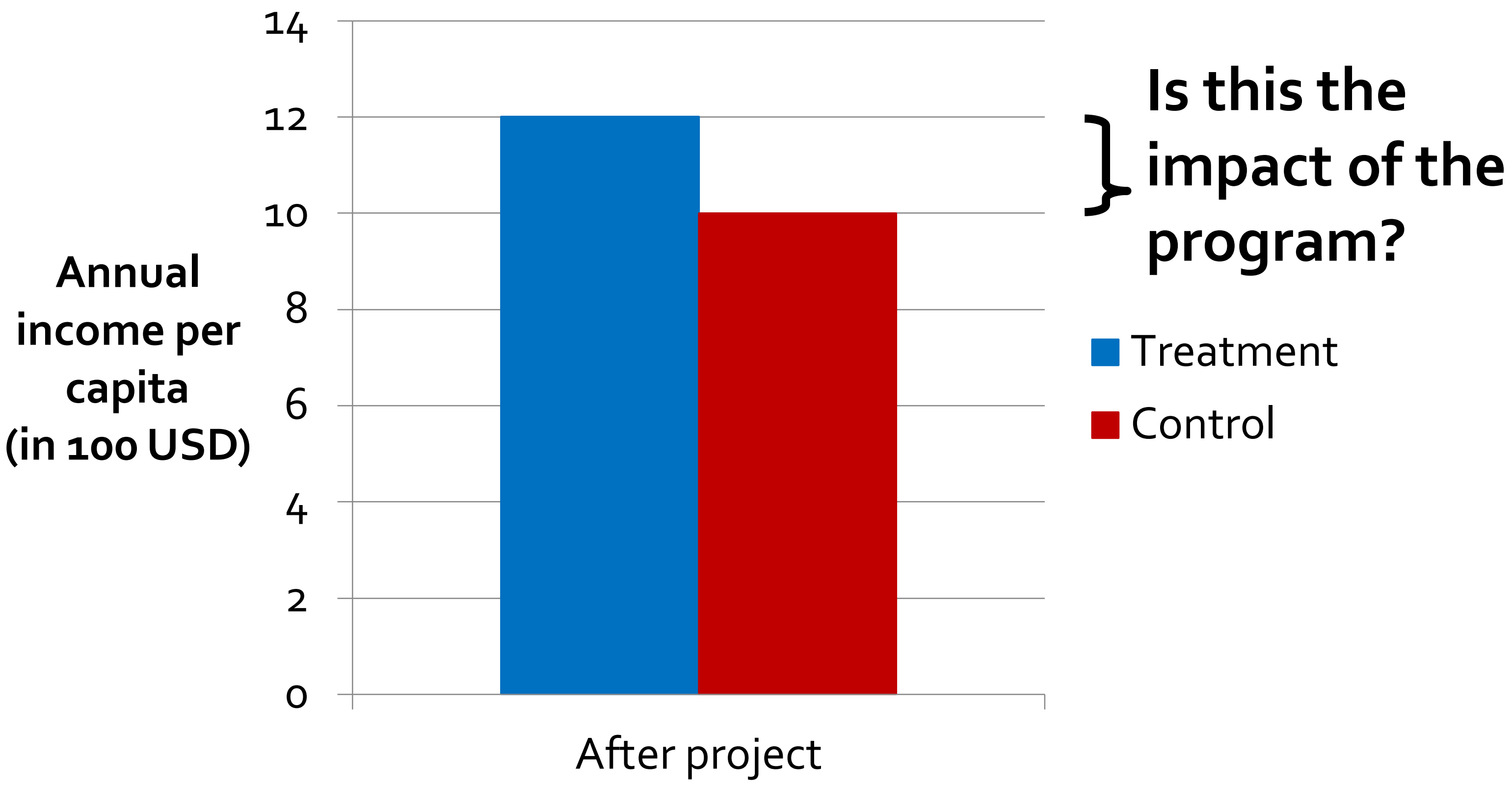
Is this the impact of the program?

Confounding factors: beard example



What if…???

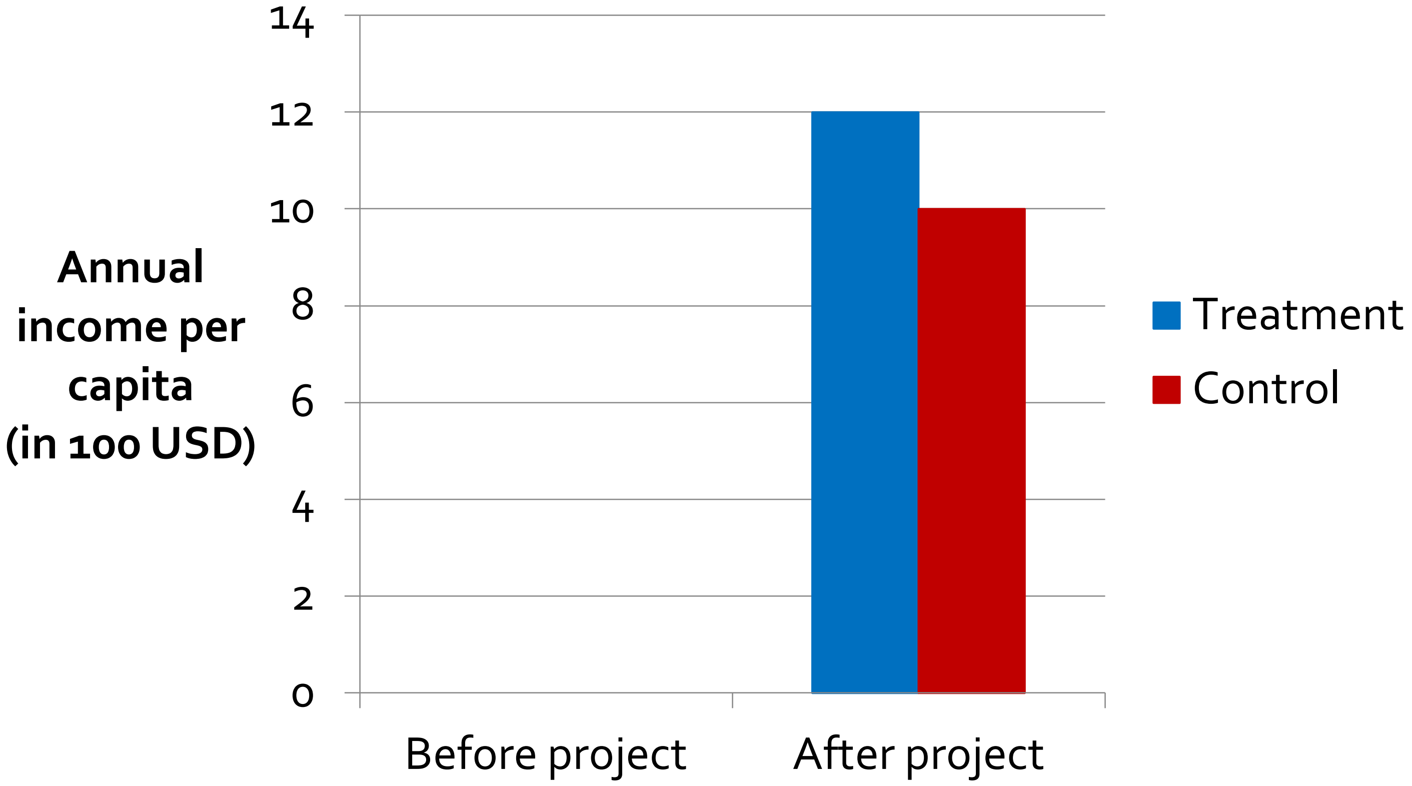
Compare treatment & control?



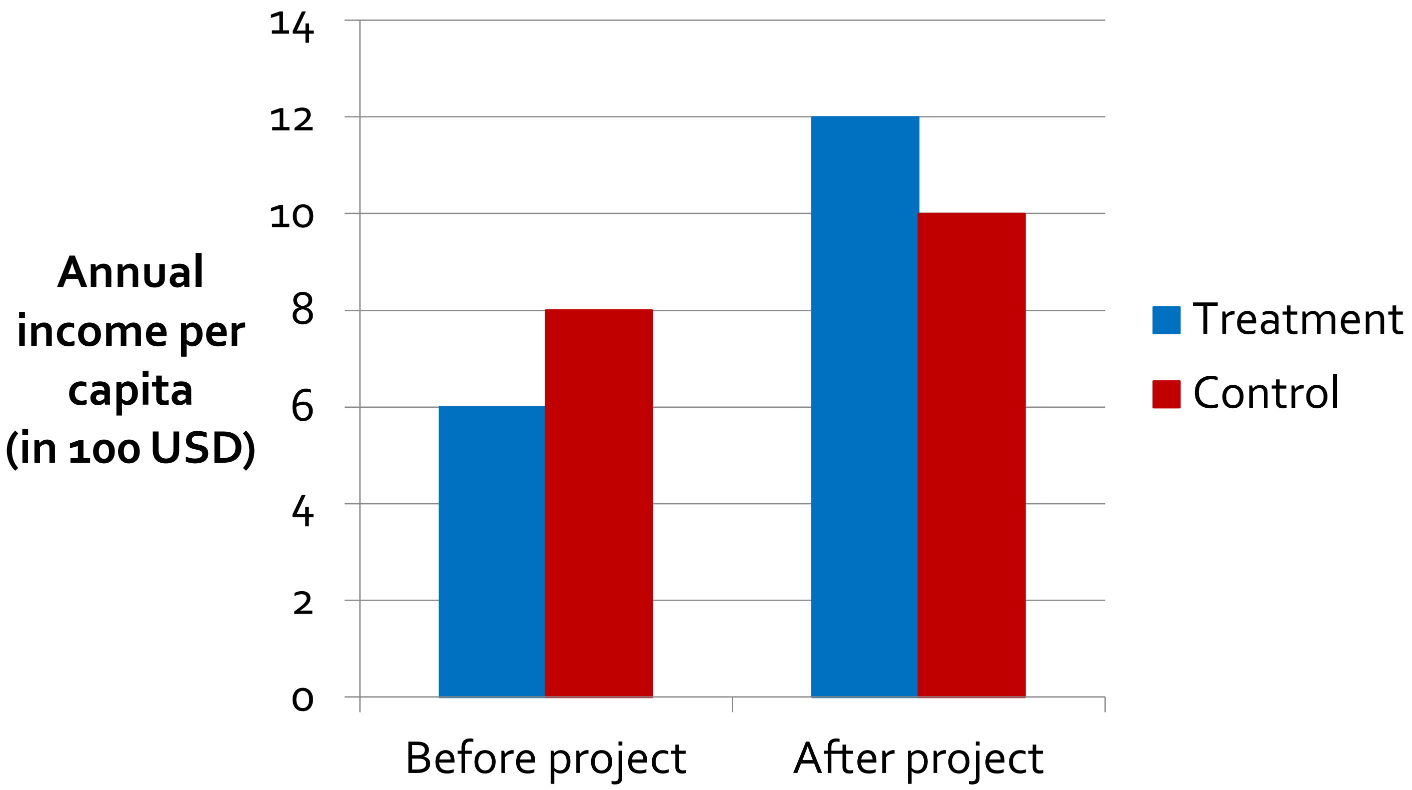
This talk

1. Difference-in-differences
2. Difference-in-differences + Matching
3. Discontinuity design

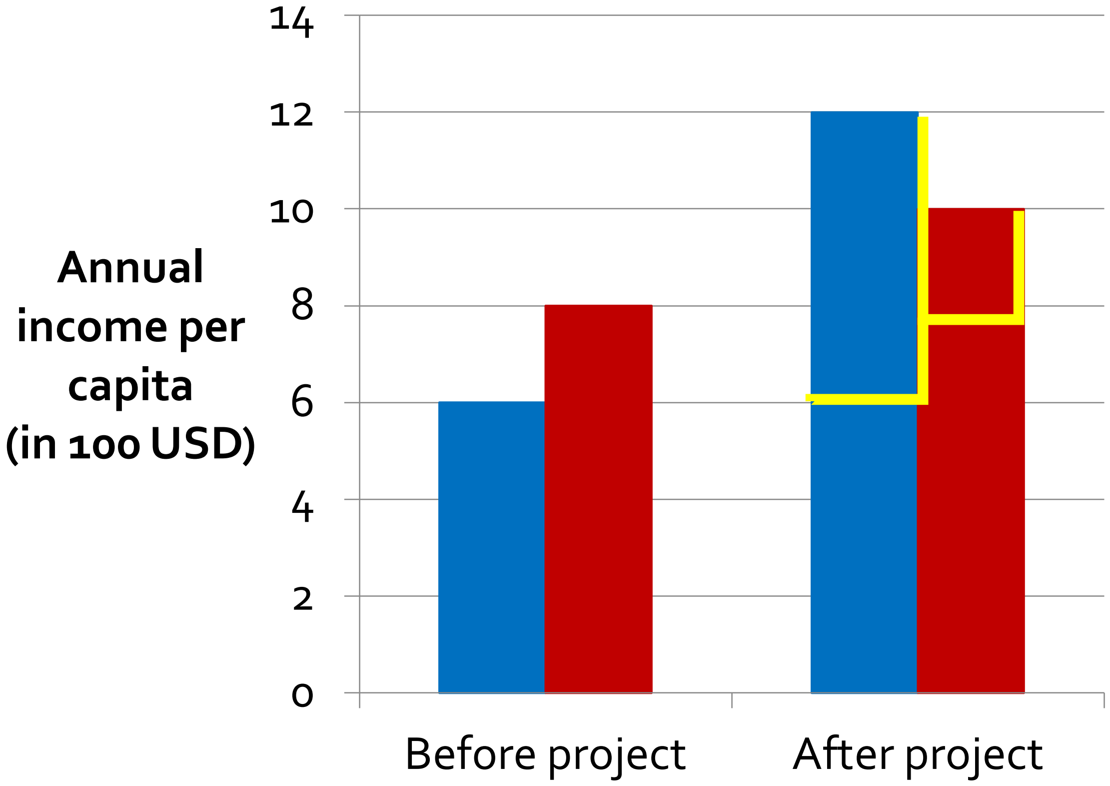
Combine the two differences!



Combine the two differences!



Difference-in-differences



Difference in treated group

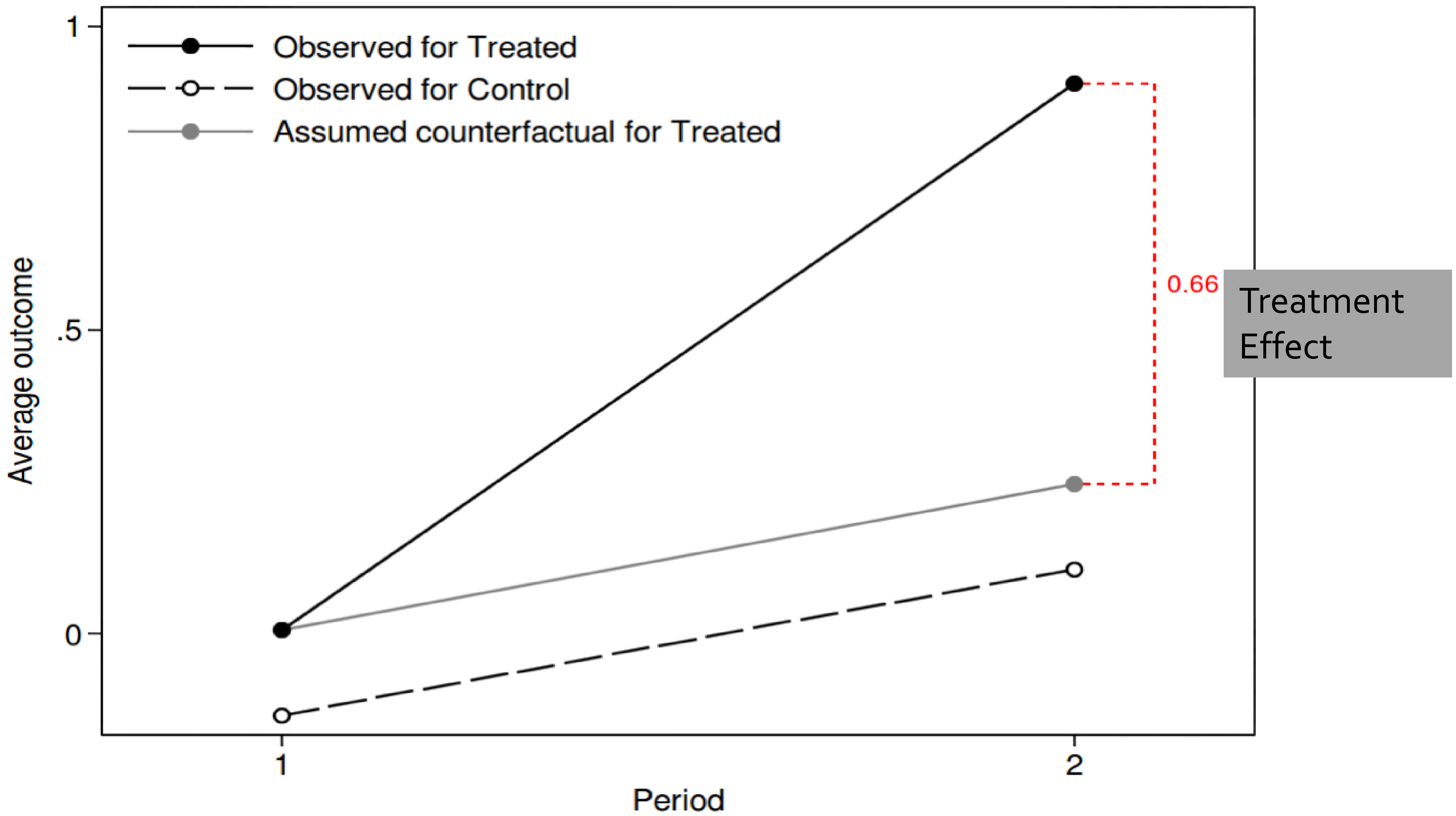
—

Difference in control group

=

Treatment effect

Key assumption: parallel trends



To make diff-in-diff work

- Can we find a plausible counterfactual?
- Try to find control group with parallel trend!

Impact of Railroad Extension in India

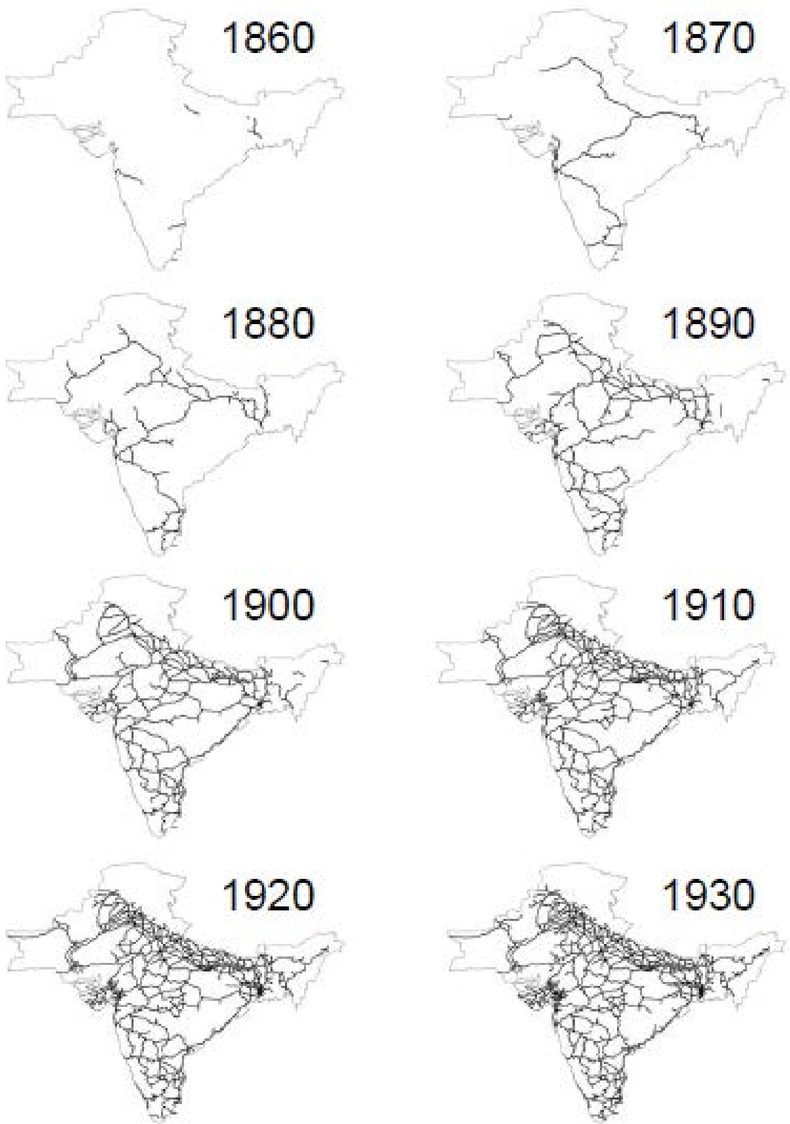
(Donaldson, American Economic Review *forthcoming*)

Intervention
Arrival of railroad network between 1870-1930

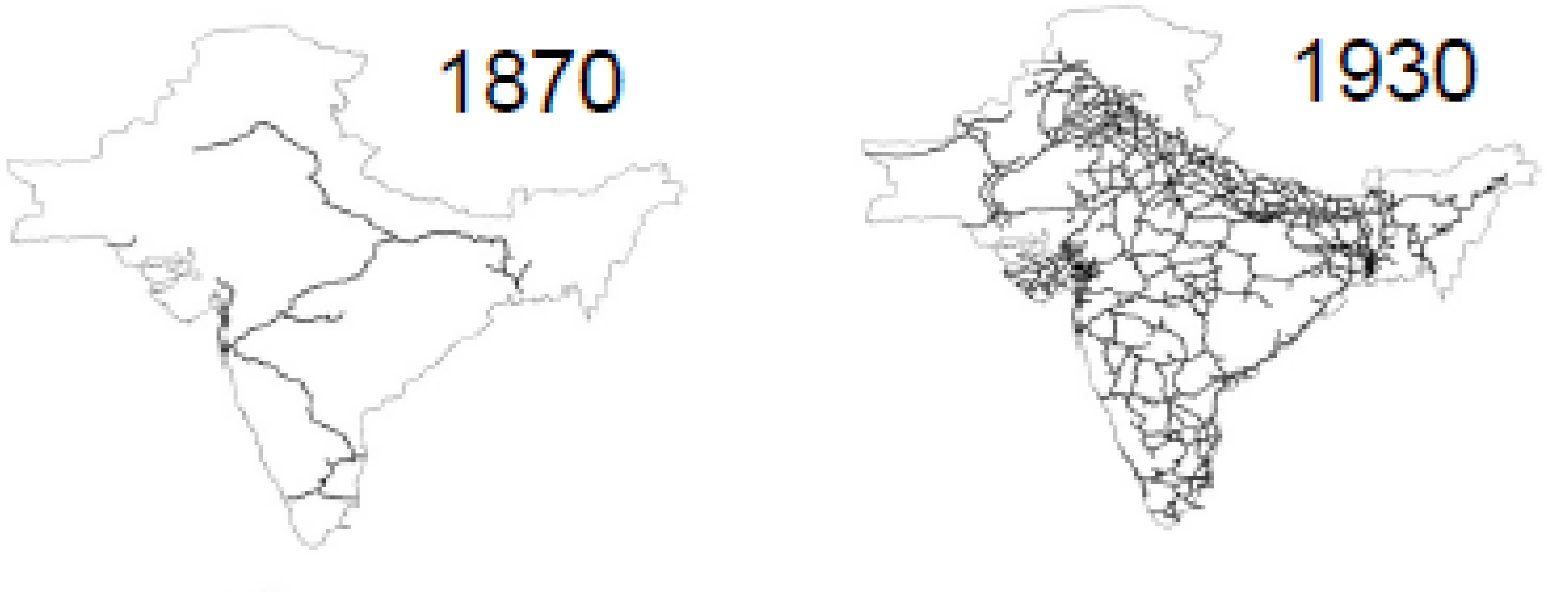
Treatment
(parts of) district is on the railroad network

Control
districts not on network

Outcome
Real agricultural income per acre



Evolution of India’s railroad network



Evolution of India’s railroad network

Impact evaluation opportunity

- New districts are connected to network every year
- Standard diff-in-diff neglects a lot of changes over time and information

Options:

- Measure impact over time (e.g., impact after 1 year differs from impact after 5 years)

- Measure a very precise average impact over time (done here through fixed effects estimation)

Are control districts a good counterfactual?

Treatment assignment related to real agricultural income?

Parallel trend violated if railroad placement decisions were based on…

– better development potential of districts or

– any other rule that relates to better development potential (e.g., commercial attractiveness of districts)

Unlikely. Military motives for railroad-building!

Possible Checks

Check parallel trend before railroad introduction

Not possible because data prior to 1870 not available

Placebo Checks

Examine relationship between real agricultural income evolution and districts considered under serious railroad expansion plans that were never realized

Very nice!!!

This talk

1. Difference-in-differences
2. Difference-in-differences + Matching
3. Discontinuity design

Diff-in-diff + Matching

- Relevant in the presence of concerns regarding parallel trend assumption
- Counterfactual: Most similar control districts
  - Each treated district is matched with a similar non-treated district based on **observable** characteristics
  - On average, matched treated districts and control districts share the same observable characteristics (by construction!)
  - Estimate the effect of program by using difference-in-differences
  - But what about unobservables??!

Impact of Rural Road Rehabilitation

(van de Walle and Cratty, World Bank Policy Research Working Paper 2007)

Intervention

Building and rehabilitation of roads at the commune level in rural Vietnam

Limitations

– Road placement rule is unclear

– One baseline survey – cannot show pre-treatment parallel trend

Possible solution

Diff-in-diff + Matching

Matching Approach

Design a control group by establishing close matches on observable characteristics

- – Geographic zone/proximity
- – similar pre-treatment income levels/living standard

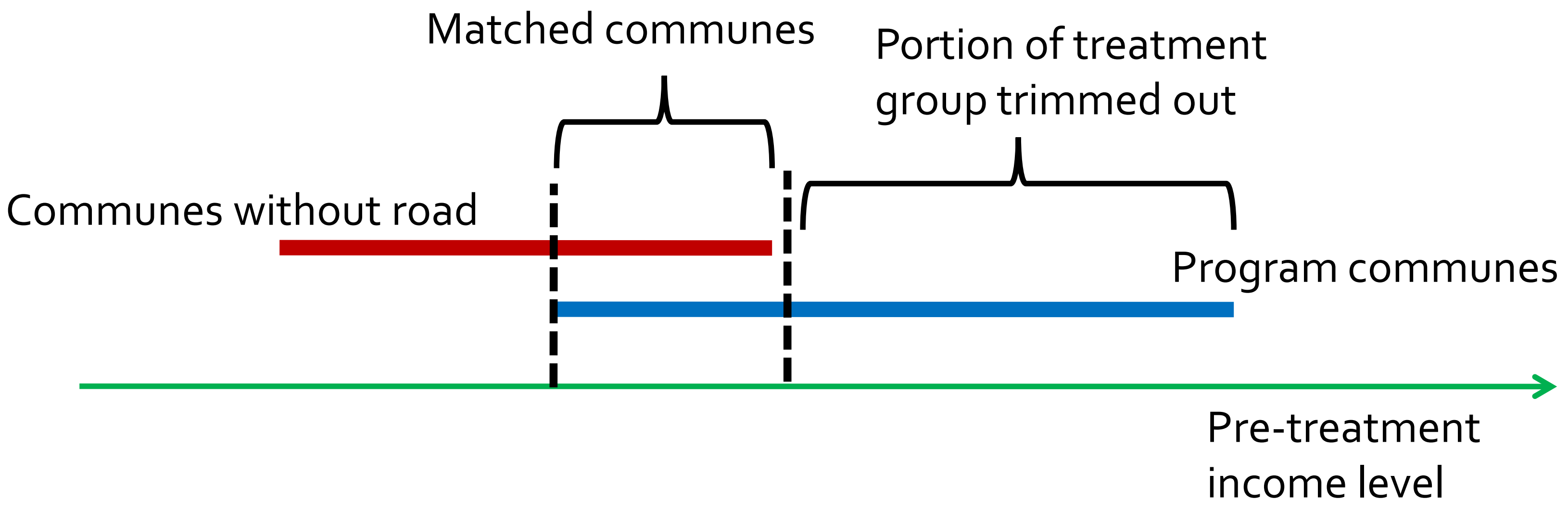
Compare only observations that have a good match

- – Treatment group: Participants that could find a match
- – Comparison group: Non-participants similar enough to the participants Compare only observations that have a good match

Implications

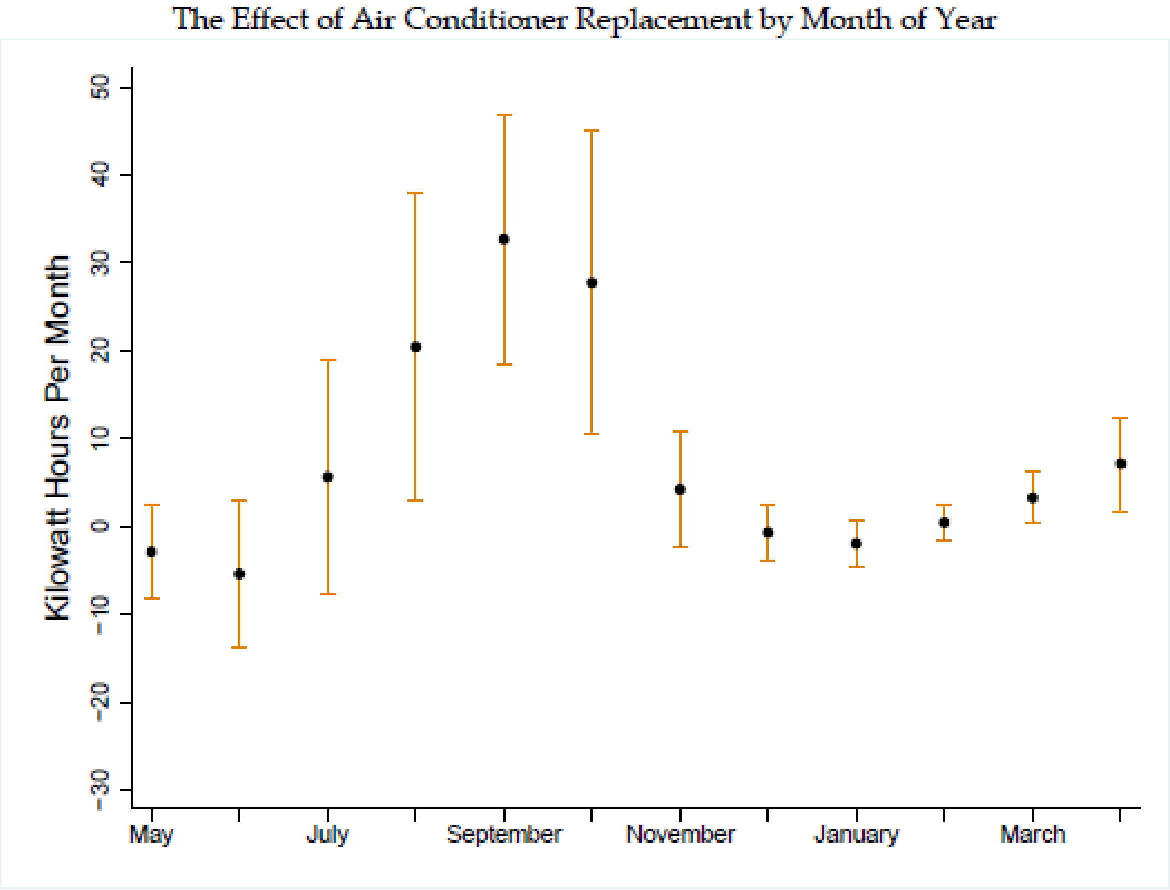
In most cases, we cannot match everyone

- Bigger sample ⟶ better matches (Costly!)
- Can't say much about the sample "trimmed out"



Matching (+ diff-in-diff): Data rich example

Cash for cooler program in Mexico



Air conditioner replacement did actually increase electricity consumption in summer

Implications

How does matching help us with our original quest for a counterfactual?

- Why did comparison households not participate?
- Could participation decision be correlated with important **unobservables**?
  - Geographic location ~~▶ weather differences?~~
  - Level of energy consumption ~~▶ pre-treatment differences~~ difference in expected energy consumption?

Conclusion

- Advantage of the matching method
  - Can help find a counterfactual where observable characteristics
- Yet
  - Hard to ignore the role of unobservable characteristics
  - We can only measure the impact for those participants that could be matched to similar non-participants
  - requires a lot of data
  - hard to predict how efficient the matching exercise will be

This talk

1. Difference-in-differences
2. Matching (+ diff-in-diff)
3. Discontinuity design

Regression discontinuity designs

- RDD is more similar to randomization
  - Based on the selection process
- Need a clear & enforced eligibility rule
  - A simple, quantifiable score
- Assignment to treatment is based on this rule
  - A threshold is established
- Compare individuals just above the threshold to individuals just below the threshold

RDD logic

- Assignment to the treatment depends on continuous “score” or ranking
  - potential beneficiaries are ordered by looking at the score
  - there is a cut-off point for “eligibility” – clearly defined criterion determined *ex ante*
  - cut-off determines the assignment to treatment
- This usually results from administrative decisions
  - resource constraints limit coverage
  - very targeted intervention
  - transparent rules

RDD in practice

- Conditional Cash Transfer program with education component in Colombia
- Poverty index score determines program eligibility
- Idea: compare program take up and school completion between…
  - Treatment group: individuals below poverty threshold
  - Comparison group: individuals above poverty threshold
- Around the threshold, **assignment to treatment is (nearly) random**
  - The only difference is program participation

RD Example 1: Poverty index score as forcing variable

Figure 1. Effects of the SISBEN Score on Participation in the Program

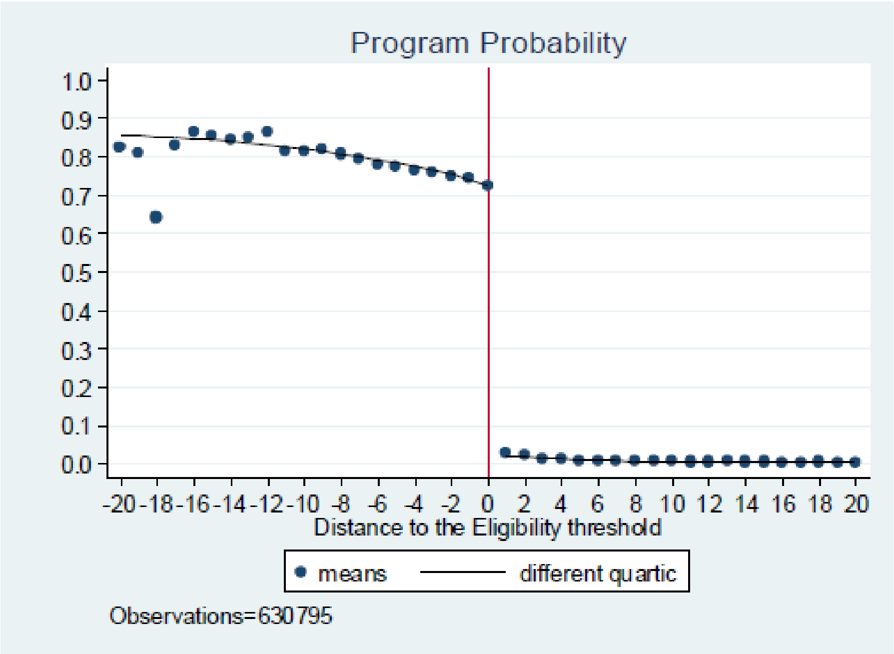
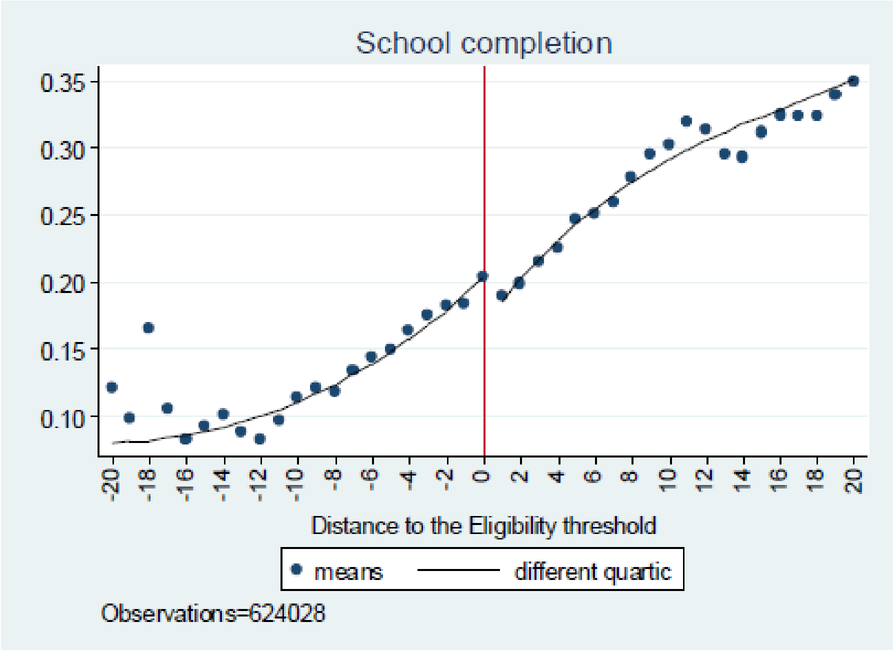
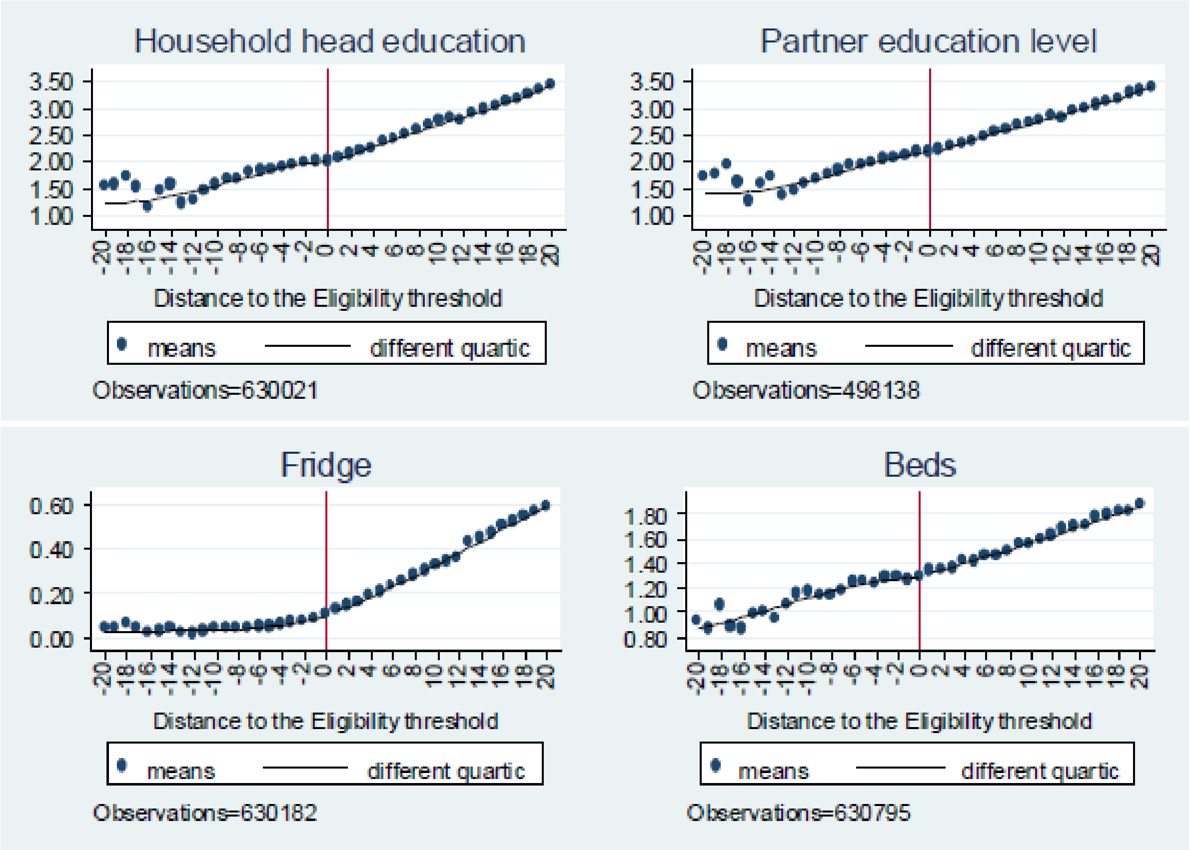


Figure 2. Impacts of FA on High School Completion (RD Analysis)



Impact of conditional cash transfers on education in Colombia
[Baez & Camacho (2011), World Bank Policy Research Working Paper 5681]

Validity assessment of RDD approach



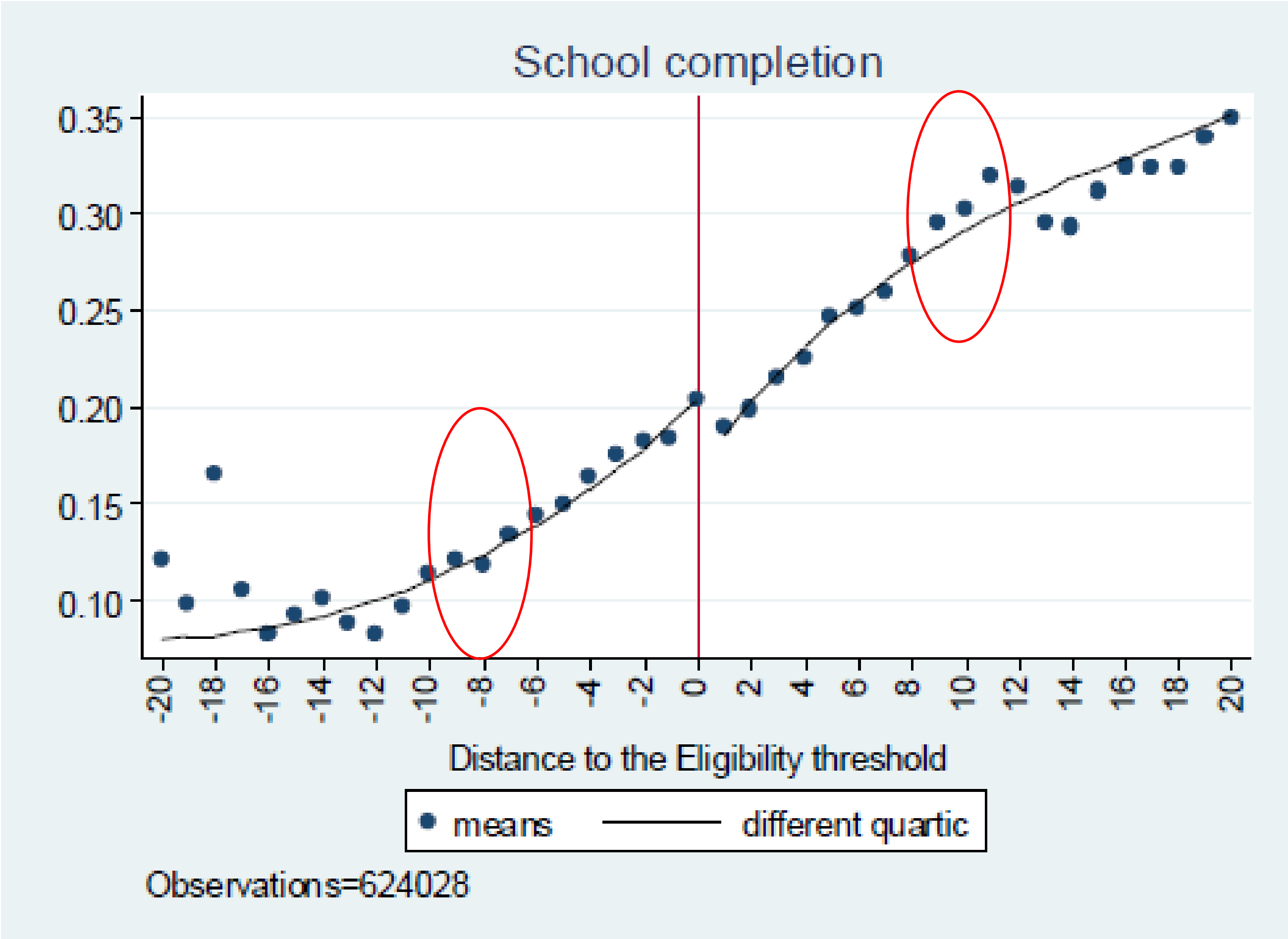
No discontinuity when using other household-level variables

RDD drawbacks

- But how generalizable are the results?
- They only tell us about how the program affects education of children in household around the poverty threshold!
- The program is likely to have different effects on the poorest

What can we learn from a very local estimate?

Figure 2. Impacts of FA on High School Completion (RD Analysis)



RD Example 2: Time as forcing variable (event study)

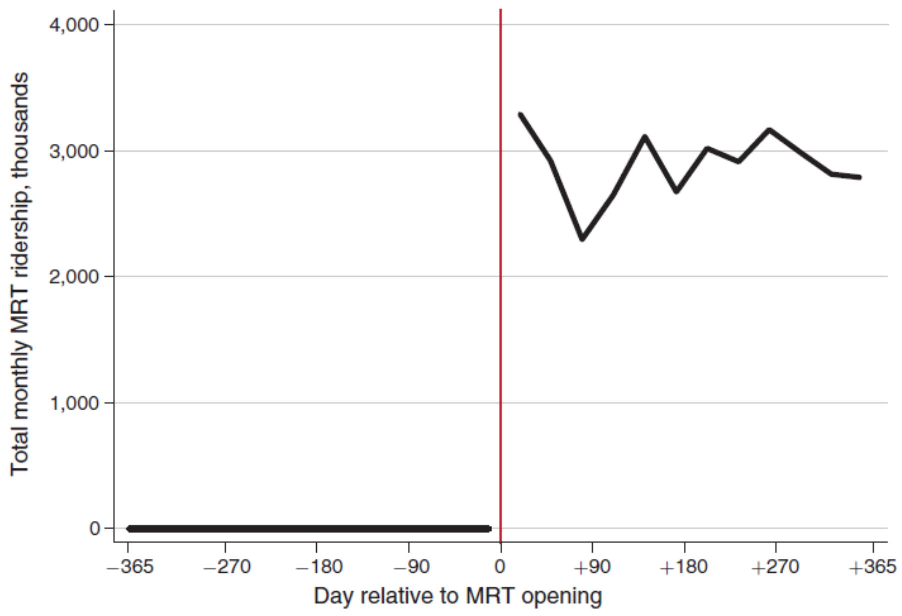
- Regression discontinuity design/event study
- Hourly air quality data
- Taipei Metro opening on 28 March 1996 at 6am
- Impact ex-ante unclear:
  - Divert marginal automobile traveler
  - Simply induce demand for travel
- Relevance: 155 million people travel on urban transit systems every day

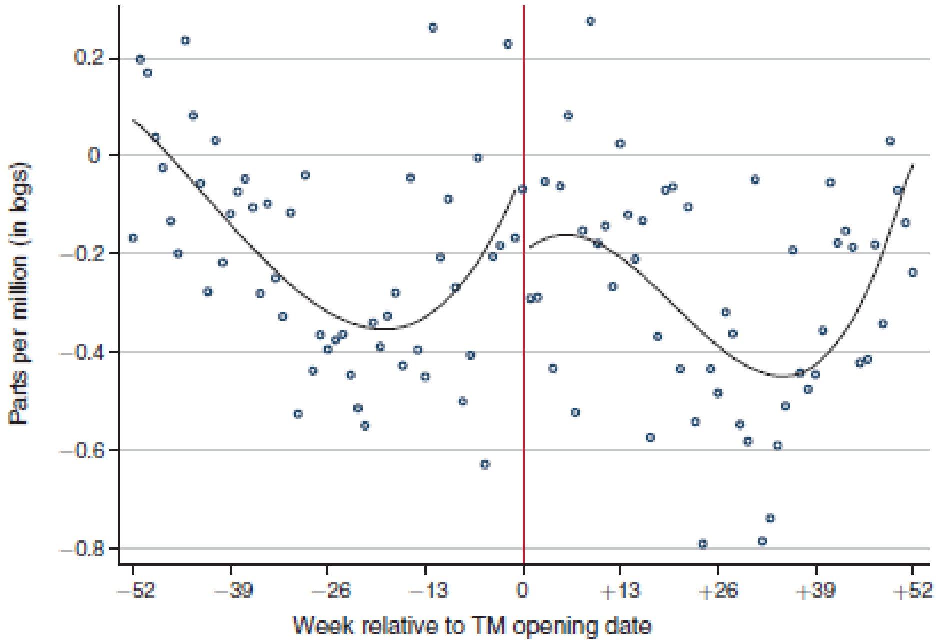
Chen and Whalley, *American Economic Journal: Economic Policy 2012*

Time as forcing variable

- Assumption: air quality would have changed smoothly in the absence of opening of the Metro
- Randomly chosen opening date (e.g., not at baseball game of national team at Summer Olympics)
- Tight oversight of the opening time-line by government regulators

RDD results





Discontinuity in metro ridership and carbon monoxide emissions (CO)
No effects on nitrogen oxides (NO_x) and ground level ozone (O_3)

Validity assessment

- No discontinuity in non-transportation source pollutants (e.g., SO_2)
- No discontinuity in Taiwan's second largest city (Kaohsiung)

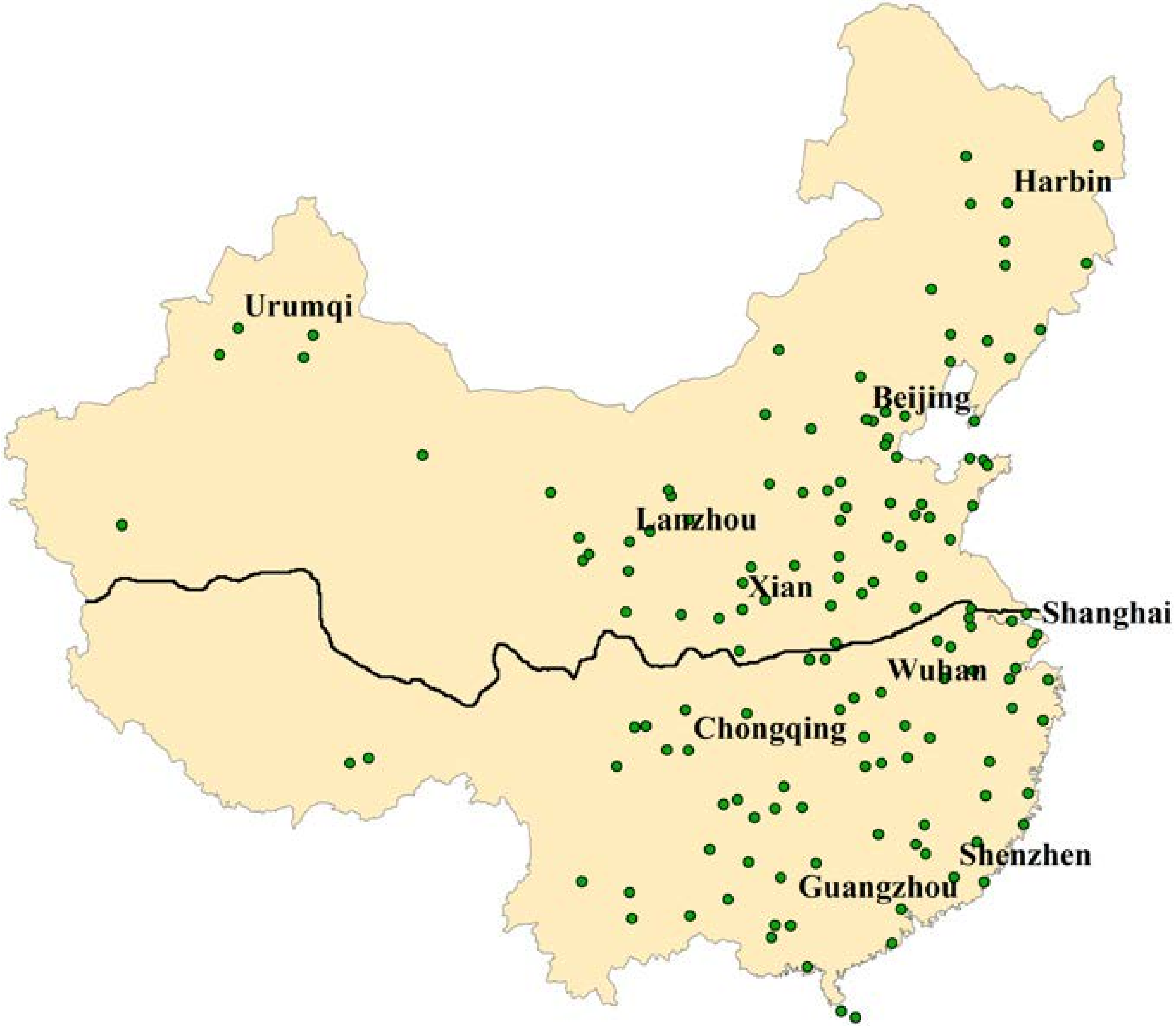
Main finding

- Direct evidence of air pollution reductions
- Indirect evidence for traffic diversion effect

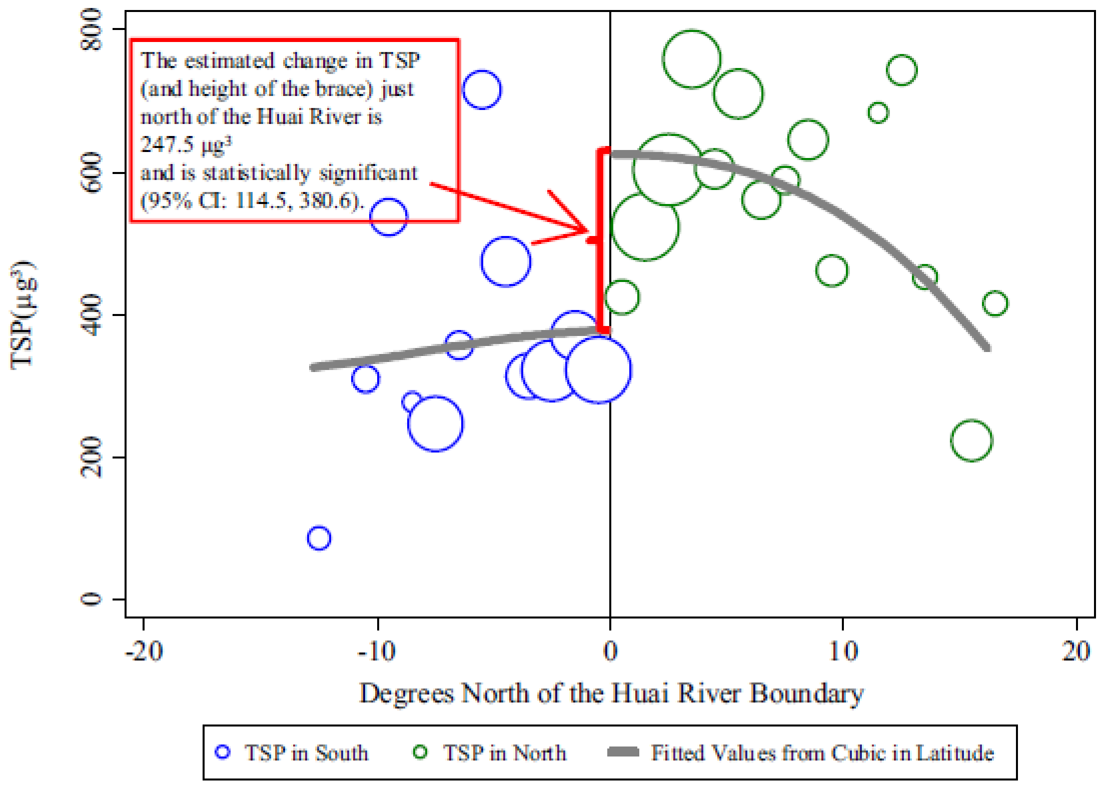
RD Example 3: Distance to geographic border as forcing variable

- Impact of sustained exposure to air pollution on life expectancy from China's Huai River policy
- Provision of free coal for heating boilers in cities north of the Huai River
- Combustion of coal in boilers is associated with the release of air pollutants (particularly particulate matter)

[Chen, Ebenstein, Greenstone, and Li (2013), *PNAS* 110 (32): 12936–12941]

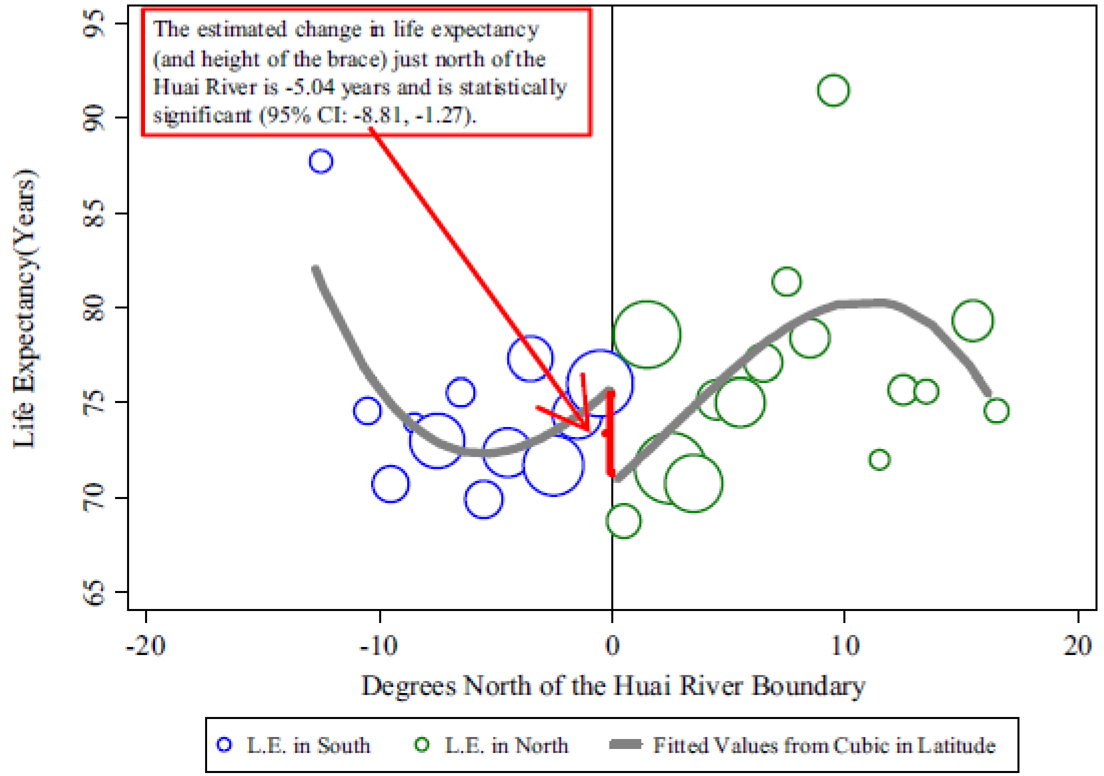

Harbin
Urumqi
Beijing
Lanzhou
Xian
Shanghai
Wuhan
Chongqing
Shenzhen
Guangzhou

Impact of policy on particulate matter



Much higher pollution North of the river

Impact of policy on life expectancy



Lower life expectancy North of the river

Validity assessment of RDD approach

- Effects on cardiorespiratory deaths but no effects on noncardiorespiratory deaths

When can we use RDDs?

- To design a prospective evaluation when randomization is not feasible
  - But need a clear allocation rule with a cut-off!
  - Poverty index and geographic borders are two examples for such an allocation rule
- To evaluate ex-post interventions using discontinuities as “natural experiments”

Summary

- Randomized controlled trials require minimal assumptions and provide intuitive estimates
- Non-experimental methods require assumptions that must be carefully tested
  - More data-intensive
  - Not always testable
- **Get creative:**
  - Mix-and-match types of methods!
  - Address relevant questions with relevant techniques