Program Evaluation with Remotely Sensed Outcomes

A discussion

Paul Goldsmith-Pinkham
Yale University
https://paulgp.com

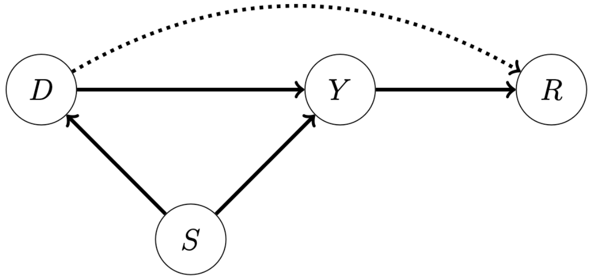
Summarize, provide alternative pitch, and then lambast

In 7 minutes I will:

1. Very briefly summarize the issue and estimator
2. Provide a connection to existing literature
3. Propose a more general frame for the paper + give examples

What is the effect of a treatment on an (unobserved) economic outcome?

- We observe two samples:
  1. Sample with treatment D and outcome R, but no outcome Y
  2. Sample with outcome Y and outcome R, but no treatment D
- D is randomly assigned in sample 1
- **But** we care about the effect of D *on* Y
  - Do not observe D and Y directly



Use the relationship between Y and R in sample 2

- We have a model between Y and R in sample 2.
  - e.g. (if R was scalar)

$$R = \beta_S Y + \varepsilon,$$

- Key **model stability** assumption: $\beta_o = \beta_e$ (in this example)
  - writ generally:

$$S \perp R \mid X, D, Y.$$

When could stability fail?

Consider the following in a geographic setting:

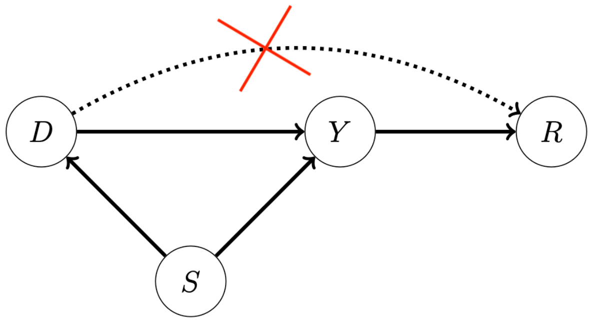
- Y is poverty in a village
- R is a satellite image of the village
- Sample e is experimental villages, sample o is observational villages

If experimental villages are in mountains, and observational villages are in plains, then stability may fail:

$\rightarrow$ terrain, vegetation, building materials lead to different images

Additional assumption of no direct effect (or ability to back out direct effect)

- D has no direct effect on R
  - **Or** observe D in sample 2 to seperate direct effect
    - Deconvolution problem
- This starts to look a lot like an IV problem, but with added twist of two samples



Two sample IV

- Other empirical settings where we don't have the relationship we're interested in the main dataset:

> Because a large data set that contains information on both age at school entry and educational attainment does not exist, we use an instrumental variables (IV) estimator that combines data derived from two independent samples.
>
> — Angrist and Krueger (1992, JASA)

- But in AK1992, the dataset has $Z[D]$ and $D[Y]$, and $Z[D]$ and $Y[R]$, and interested in $D[Y]$ on $Y[R]$
  - Here, we have D and R, and Y and R. Not the same!

Paper's Estimator w/o covariates and scalar R

$$\theta = \frac{\mathbb{E}\{R'\Delta^e\}}{\mathbb{E}\{R'\Delta^o\}},$$

where

$$\Delta^e := \underbrace{\frac{D\mathbb{1}\{S=e\}}{p_d^e} - \frac{(1-D)\mathbb{1}\{S=e\}}{1-p_d^e}}_{\text{inverse-propensity-weighted treatment indicator}}$$

and

$$\Delta^o := \underbrace{\frac{Y\mathbb{1}\{S=o\}}{p_y^o} - \frac{(1-Y)\mathbb{1}\{S=o\}}{1-p_y^o}}_{\text{inverse-probability-weighted outcome indicator}}$$

where $p_d^e = \Pr(D=1, S=e)$ and $p_y^o = \Pr(Y=1, S=o)$.

Estimator intuition

- Δ^e varies at different levels of R
- Δ^o varies at different levels of R
- Ratio of projections recovers effect of D on Y
  - Variation in R predicts changes in both Δ^e and Δ^o
  - Under stability, changes in Δ^e map to changes in Δ^o
- R is a *noisy* measure for Y that helps us identify causal relationship

$$\Delta^e = \theta \Delta^o + u$$

Alternative framing: BPP/Hall-Mishkin consumption models

- Permanent income shocks ζ_t (unobserved)
- Transitory income shocks ε_t (unobserved)
- Observed:
  - income y_t, $\Delta y_t = \zeta_t + \varepsilon_t$
  - consumption c_t (function of permanent income by PIH...)

Estimator:

$$\varphi = \frac{\mathrm{Cov}(\Delta c_t, \Delta y_{t+1})}{\mathrm{Cov}(\Delta y_t, \Delta y_{t+1})}$$

Use **future** income changes to isolate permanent income shocks in today's income

Alternative framing: BPP/Hall-Mishkin consumption models

- Similar to RSV, BPP/HM have many testable implications (since you can also use multiple future periods' income as instruments)
- Identification comes from timing restriction in BPP/HM
  - Similar to stability? (relationship between income across time not changing)
- Prompts an interesting question about **timing** in this paper
  - Can we think more about dynamics and timing?
  - Can we think about more applications?

Is this paper more general?

- Conclude with more examples where I think this holds beyond remote sensing

1. Criminal Justice / Recidivism
    - D = treatment (e.g. job training)
    - Y = actual criminal behavior (unobserved)
    - R = proxy for criminal behavior (e.g. arrests, police stops)

- Many reentry experiments only track arrests through administrative records
- Arrests are a consequence of criminal activity, not the activity itself
- An auxiliary dataset could link self-reported criminal behavior surveys to arrest records

Is this paper more general?

2. Education
   - D = Teaching intervention (e.g. new tool)
   - Y = Deep conceptual understanding (requires expert assessment)
   - R = Standardized test scores

- Large-scale education RCTs often can only afford to collect test scores
- Auxiliary data could come from cognitive science studies that administer both detailed assessments and standardized tests

Is this paper more general?

3. Clinical Trials

- D = Drug or behavioral intervention
- Y = Underlying disease state or health status (requires expensive clinical workup)
- R = **Downstream** symptom or cheap biomarker that is caused by the health state

Is this paper more general?

4. Social Networks
  - D = Edge building intervention (e.g. friend recommendation)
  - Y = True social connectedness or relationship quality (requires detailed surveys)
  - R = Observed communication patterns—calls, texts, co-location (downstream of relationships)

- Communication data is cheap (call records, app data)
- Communication is caused by relationships
- Can be high dimensional!

Many others...

- Fun paper, but I think it could be more ambitious!
- Really important that they nail the framing of issue of using surrogacy approach instead of this approach
  - Clear bias issues!
- Lots of other applications beyond remote sensing