Program Evaluation with Remotely Sensed Outcomes

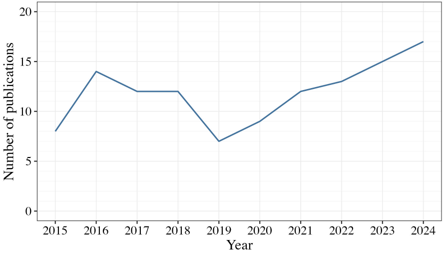
Ashesh Rambachan[1], Rahul Singh[2], Davide Viviano[2]

[1]MIT Economics
[2]Harvard Economics

NBER SI 2025

Economists increasingly use remote sensing

- a general trend in top economics journals



- often in environmental and development economics

Why is remote sensing necessary?

- economic outcomes may be costly or infeasible to collect
  - environmental quality, e.g. crop burning
  - living standards, e.g. household consumption

- so researchers use remotely sensed variables (RSVs)
  - satellite images (Jean et al. 2016, Jayachandran et al. 2017, Aiken et al. 2022, Currie et al. 2023, Balboni et al. 2024, Jack et al. 2025)
  - night lights (Chen + Nordhaus 2011, Henderson et al. 2012, Asher et al. 2021)
  - roofing material (Marx et al. 2019, Michaels et al. 2021, Huang et al. 2021)

- this paper: program evaluation with remotely sensed outcomes?

Why is remote sensing necessary?

- economic outcomes may be costly or infeasible to collect
  - environmental quality, e.g. crop burning
  - living standards, e.g. household consumption
- so researchers use remotely sensed variables (RSVs)
  - satellite images (Jean et al. 2016, Jayachandran et al. 2017, Aiken et al. 2022, Currie et al. 2023, Balboni et al. 2024, Jack et al. 2025)
  - night lights (Chen + Nordhaus 2011, Henderson et al. 2012, Asher et al. 2021)
  - roofing material (Marx et al. 2019, Michaels et al. 2021, Huang et al. 2021)
- this paper: program evaluation with remotely sensed outcomes?

Why is remote sensing necessary?

- economic outcomes may be costly or infeasible to collect
  - environmental quality, e.g. crop burning
  - living standards, e.g. household consumption

- so researchers use remotely sensed variables (RSVs)
  - satellite images (Jean et al. 2016, Jayachandran et al. 2017, Aiken et al. 2022, Currie et al. 2023, Balboni et al. 2024, Jack et al. 2025)
  - night lights (Chen + Nordhaus 2011, Henderson et al. 2012, Asher et al. 2021)
  - roofing material (Marx et al. 2019, Michaels et al. 2021, Huang et al. 2021)

- this paper: program evaluation with remotely sensed outcomes?

Main idea

- suppose the researcher has two data sets
  1. experimental: treatment, RSV
  2. observational: outcome, RSV

- we study the RSV as a post-outcome variable
  - e.g. fires cause changes in satellite images; not vice versa

- we propose a new method
  - comparing predicted outcomes of treated, untreated has bias
  - novel formula to identify treatment effect by data combination
  - for efficiency, conduct three predictions rather than one

Main idea

- suppose the researcher has two data sets
  1. experimental: treatment, RSV
  2. observational: outcome, RSV

- we study the RSV as a post-outcome variable
  - e.g. fires cause changes in satellite images; not vice versa

- we propose a new method
  - comparing predicted outcomes of treated, untreated has bias
  - novel formula to identify treatment effect by data combination
  - for efficiency, conduct three predictions rather than one

Main idea

- suppose the researcher has two data sets
  1. experimental: treatment, RSV
  2. observational: outcome, RSV
- we study the RSV as a post-outcome variable
  - e.g. fires cause changes in satellite images; not vice versa
- we propose a new method
  - comparing predicted outcomes of treated, untreated has bias
  - novel formula to identify treatment effect by data combination
  - for efficiency, conduct three predictions rather than one

Related work

- auxiliary variable models in causal inference
  - surrogates are *pre*-outcome variables (Athey et al. 2024, Kallus + Mao 2024)
  - misusing an RSV as a surrogate leads to arbitrary biases
- prediction powered inference
  - machine learning predictors as surrogates (Angelopoulos et al. 2023, Lu et al. 2025, Kluger et al. 2025, Ji et al 2025)
- generative models
  - must be correctly specified (Gentzkow et al. 2019, Alix-Garcia + Millimet 2023, Proctor et al. 2023, Battaglia et al. 2024)
- data combination
  - highly general (Cross + Manski 2002, Chen et al. 2005 + 2008, Ridder + Moffitt 2007, Bareinboim + Pearl 2016, Graham et al. 2016, D'Haultfoeuille et al. 2025)
  - opposite assumption: stability of outcome, versus stability of RSV

Motivation: Takeaway

program evaluation from *post*-outcome variable?

Outline

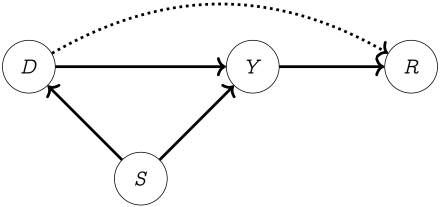
1 Model

2 Common practice

3 Proposal

4 Case study

Model: Two samples



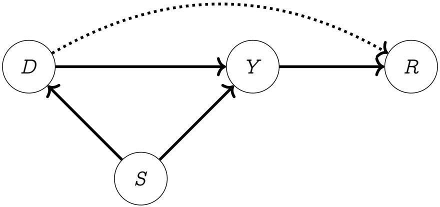
- $S \in \{e, o\}$ sample indicator
- $D \in \{0, 1\}$ treatment
- $Y \in \{0, 1\}$ outcome (discrete or continuous in the paper)
- $R \in \mathcal{R}$ remotely sensed outcome
- $X \in \mathcal{X}$ covariate (see paper)

- goal: treatment effect in the experiment

$$\theta_0 = \mathbb{E}\{Y(1) - Y(0) | S = e\}$$

- there are two imperfect samples (extensions in the paper)
  1. experimental $(S = e)$: D, R
  2. observational $(S = o)$: Y, R, and possibly D

Model: Two samples



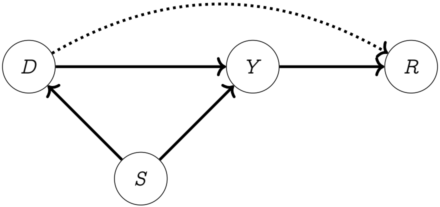
- $S \in \{e, o\}$ sample indicator
- $D \in \{0, 1\}$ treatment
- $Y \in \{0, 1\}$ outcome (discrete or continuous in the paper)
- $R \in \mathcal{R}$ remotely sensed outcome
- $X \in \mathcal{X}$ covariate (see paper)

- goal: treatment effect in the experiment

$$\theta_0 = \mathbb{E}\{Y(1) - Y(0) | S = e\}$$

- there are two imperfect samples (extensions in the paper)
  1. experimental ($S = e$): D, R
  2. observational ($S = o$): Y, R, and possibly D

Model: Two samples



- $S \in \{e, o\}$ sample indicator
- $D \in \{0, 1\}$ treatment
- $Y \in \{0, 1\}$ outcome (discrete or continuous in the paper)
- $R \in \mathcal{R}$ remotely sensed outcome
- $X \in \mathcal{X}$ covariate (see paper)

- goal: treatment effect in the experiment

$$\theta_0 = \mathbb{E}\{Y(1) - Y(0) | S = e\}$$

- there are two imperfect samples (extensions in the paper)
  1. experimental ($S = e$): D, R
  2. observational ($S = o$): Y, R, and possibly D

Model: Assumptions

We place three assumptions

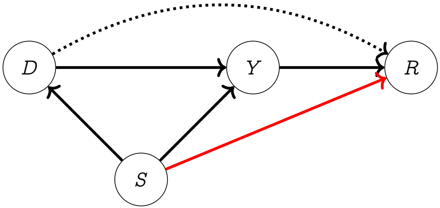
1. The experimental sample was properly collected
   - D was randomly assigned to units: $D \perp\!\!\!\perp \{Y(1), Y(0)\} | S = e$
   - this assumption is satisfied by design
2. The distribution of the RSV is stable across samples
3. The observational sample is "complete"

Model: Assumptions

We place three assumptions

1. The experimental sample was properly collected

2. The distribution of the RSV is stable across samples



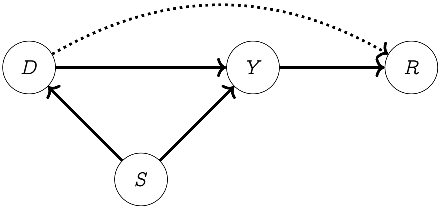
- our main assumption: $S \perp\!\!\!\perp R | D, Y$
- assess by diagnostic tests and plots (later in this talk)

3. The observational sample is "complete"

Model: Assumptions

We place three assumptions

1. The experimental sample was properly collected
2. The distribution of R is stable across samples
3. The observational sample is "complete":



(i) either we observe D in the observational sample;
(ii) or no dashed line: $D \perp\!\!\!\perp R | Y$ (for today)

Model: Takeaway

RSV is *stable* across samples

Outline

1 Model

2 Common practice

3 Proposal

4 Case study

Common practice: Method

What is a common practice?

1. train a predictor with obs. sample: $f(R) = \mathbb{E}(Y|R, S = o)$
2. compare predicted outcome of treated, untreated in exp. sample

$$\tilde{\theta} = \mathbb{E}\{f(R)|D = 1, S = e\} - \mathbb{E}\{f(R)|D = 0, S = e\}$$

Interpretation

- in $\sim 50\%$ of general interest papers with remotely sensed outcomes
- implicitly uses the RSV as a pre-outcome *surrogate* (Athey et al. 2024)

(a) Our model: Post-outcome (b) Surrogate model: Pre-outcome

Common practice: Method

What is a common practice?

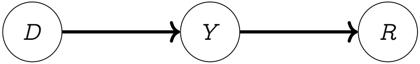
1 train a predictor with obs. sample: $f(R) = \mathbb{E}(Y|R, S = o)$

2 compare predicted outcome of treated, untreated in exp. sample

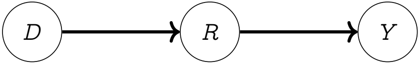
$$\tilde{\theta} = \mathbb{E}\{f(R)|D = 1, S = e\} - \mathbb{E}\{f(R)|D = 0, S = e\}$$

Interpretation

- in $\sim 50\%$ of general interest papers with remotely sensed outcomes
- implicitly uses the RSV as a pre-outcome *surrogate* (Athey et al. 2024)



(a) Our model: Post-outcome



(b) Surrogate model: Pre-outcome

Common practice: Bias

What goes wrong? Recall $f(R) = \mathbb{E}(Y|R, S = o)$

$$\tilde{\theta} = \mathbb{E}\{f(R)|D = 1, S = e\} - \mathbb{E}\{f(R)|D = 0, S = e\}$$

Three levels of critique

1. warm up: bias in the "irrelevance" case
   - suppose the RSV fails to predict the outcome
   - then $f(R)$ is constant and $\tilde{\theta} = 0$!
2. bias in the linear case
   - suppose $f(R) = \tilde{\beta}_0 + \tilde{\beta} R$ and $\mathbb{E}(R|Y, D, S = e) = \beta_0 + \beta Y$
   - combining these expressions, $\tilde{\theta} = \tilde{\beta}\beta\theta_0$
3. bias in general
   - Proposition (informal): bias of $\tilde{\theta}$ can be positive or negative
   - significant in practice: underestimate effect on crop burning by $\sim 47\%$

Common practice: Bias

What goes wrong? Recall $f(R) = \mathbb{E}(Y|R, S = o)$

$$\tilde{\theta} = \mathbb{E}\{f(R)|D = 1, S = e\} - \mathbb{E}\{f(R)|D = 0, S = e\}$$

Three levels of critique

1. warm up: bias in the "irrelevance" case
   - suppose the RSV fails to predict the outcome
   - then $f(R)$ is constant and $\tilde{\theta} = 0$!
2. bias in the linear case
   - suppose $f(R) = \tilde{\beta}_0 + \tilde{\beta} R$ and $\mathbb{E}(R|Y, D, S = e) = \beta_0 + \beta Y$
   - combining these expressions, $\tilde{\theta} = \tilde{\beta}\beta\theta_0$
3. bias in general
   - Proposition (informal): bias of $\tilde{\theta}$ can be positive or negative
   - significant in practice: underestimate effect on crop burning by $\sim 47\%$

Common practice: Bias

What goes wrong? Recall $f(R) = \mathbb{E}(Y|R, S = o)$

$$\tilde{\theta} = \mathbb{E}\{f(R)|D = 1, S = e\} - \mathbb{E}\{f(R)|D = 0, S = e\}$$

Three levels of critique

1. warm up: bias in the "irrelevance" case
   - suppose the RSV fails to predict the outcome
   - then $f(R)$ is constant and $\tilde{\theta} = 0$!
2. bias in the linear case
   - suppose $f(R) = \tilde{\beta}_0 + \tilde{\beta}R$ and $\mathbb{E}(R|Y, D, S = e) = \beta_0 + \beta Y$
   - combining these expressions, $\tilde{\theta} = \tilde{\beta}\beta\theta_0$
3. bias in general
   - Proposition (informal): bias of $\tilde{\theta}$ can be positive or negative
   - significant in practice: underestimate effect on crop burning by $\sim 47\%$

Common practice: Takeaway

what should we do instead?

Outline

1 Model

2 Common practice

3 Proposal

4 Case study

Proposal: Identification

Our main contribution is identification, i.e. an RSV formula.

Define the treatment variation $\Delta^e = \frac{1(D=1,S=e)}{\mathbb{P}(D=1,S=e)} - \frac{1(D=0,S=e)}{\mathbb{P}(D=0,S=e)}$.

Define the outcome variation $\Delta^o = \frac{1(Y=1,S=o)}{\mathbb{P}(Y=1,S=o)} - \frac{1(Y=0,S=o)}{\mathbb{P}(Y=0,S=o)}$.

Theorem (informal): Under Assumptions 1, 2, and 3(ii), $\theta_0 = \frac{\mathbb{E}(\Delta^e|R)}{\mathbb{E}(\Delta^o|R)}$.

Corollary (informal): For any representation $H(R)$, $\theta_0 = \frac{\mathbb{E}\{\Delta^e H(R)\}}{\mathbb{E}\{\Delta^o H(R)\}}$.

Proposal: Identification

Our main contribution is identification, i.e. an RSV formula.

Define the treatment variation $\Delta^e = \frac{1(D=1,S=e)}{\mathbb{P}(D=1,S=e)} - \frac{1(D=0,S=e)}{\mathbb{P}(D=0,S=e)}$.

Define the outcome variation $\Delta^o = \frac{1(Y=1,S=o)}{\mathbb{P}(Y=1,S=o)} - \frac{1(Y=0,S=o)}{\mathbb{P}(Y=0,S=o)}$.

Theorem (informal): Under Assumptions 1, 2, and 3(ii), $\theta_0 = \frac{\mathbb{E}(\Delta^e|R)}{\mathbb{E}(\Delta^o|R)}$.

Corollary (informal): For any representation $H(R)$, $\theta_0 = \frac{\mathbb{E}\{\Delta^e H(R)\}}{\mathbb{E}\{\Delta^o H(R)\}}$.

Proposal: Identification

Our main contribution is identification, i.e. an RSV formula.

Define the treatment variation $\Delta^e = \frac{1(D=1,S=e)}{\mathbb{P}(D=1,S=e)} - \frac{1(D=0,S=e)}{\mathbb{P}(D=0,S=e)}$.

Define the outcome variation $\Delta^o = \frac{1(Y=1,S=o)}{\mathbb{P}(Y=1,S=o)} - \frac{1(Y=0,S=o)}{\mathbb{P}(Y=0,S=o)}$.

Theorem (informal): Under Assumptions 1, 2, and 3(ii), $\theta_0 = \frac{\mathbb{E}(\Delta^e|R)}{\mathbb{E}(\Delta^o|R)}$.

Corollary (informal): For any representation $H(R)$, $\theta_0 = \frac{\mathbb{E}\{\Delta^e H(R)\}}{\mathbb{E}\{\Delta^o H(R)\}}$.

Proposal: Identification

Our main contribution is identification, i.e. an RSV formula.

Define the treatment variation $\Delta^e = \frac{1(D=1,S=e)}{\mathbb{P}(D=1,S=e)} - \frac{1(D=0,S=e)}{\mathbb{P}(D=0,S=e)}$.

Define the outcome variation $\Delta^o = \frac{1(Y=1,S=o)}{\mathbb{P}(Y=1,S=o)} - \frac{1(Y=0,S=o)}{\mathbb{P}(Y=0,S=o)}$.

Theorem (informal): Under Assumptions 1, 2, and 3(ii), $\theta_0 = \frac{\mathbb{E}(\Delta^e | R)}{\mathbb{E}(\Delta^o | R)}$.

Corollary (informal): For any representation $H(R)$, $\theta_0 = \frac{\mathbb{E}\{\Delta^e H(R)\}}{\mathbb{E}\{\Delta^o H(R)\}}$.

Proposal: Identification

Let's interpret

$$\theta_0 = \frac{\mathbb{E}(\Delta^e | R)}{\mathbb{E}(\Delta^o | R)}, \quad \theta_0 = \frac{\mathbb{E}\{\Delta^e H(R)\}}{\mathbb{E}\{\Delta^o H(R)\}}.$$

- numerator and denominator from different samples
  - numerator is effect of D and Y on R
  - so divide by effect of Y on R
- no need to specify the distribution of $R|Y$
- any representation $H(R)$ is valid
  - as long as the representation is predictive: $\mathbb{E}\{\Delta^o H(R)\} \neq 0$
  - weak RSV test: is $\mathbb{E}\{\Delta^o H(R)\} \approx 0$?
  - joint test: do $H(R)$ and $H'(R)$ give similar estimates?

Proposal: Efficient inference

Which representation is optimal?

Corollary (informal): Efficiency when $H^*(R) = \frac{\mathbb{E}(\Delta^o | R)}{\mathbb{E}\{(\Delta^e - \Delta^o \theta_0)^2 | R\}}$.

Interpretation

- from theory of optimal instruments (Chamberlain 1987, Newey 1993)
- three predictions rather than one
  1. outcome from RSV
  2. treatment from RSV
  3. sample indicator from RSV
- inference by cross fitting (Angrist et al. 1999, Chernozhukov et al. 2018 + 2023)
  - learn optimal representation (via three predictions) on one fold
  - estimate treatment effect on other fold
  - valid using any mis-specified ML (Mackey et al. 2018, Chen et al. 2020)

Proposal: Takeaway

a novel formula for data combination

Outline

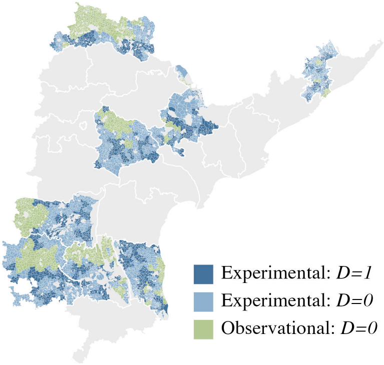
1 Model

2 Common practice

3 Proposal

4 Case study

Case study: Smartcards

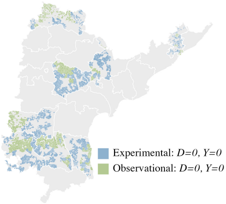


- what is the effect of Smartcards on poverty in Andhra Pradesh?
- we merge a real randomized experiment (Muralidharan et al. 2023) with real satellite images (Asher et al 2021, Rolf et al. 2021)
- we create two imperfect samples
  1. experimental: Smartcard status, satellite image
  2. observational: poverty level, satellite image

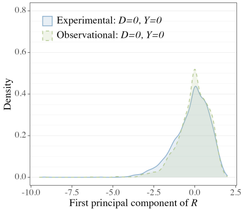
Case study: Key assumption

Our key assumption is plausible: $S \perp\!\!\!\perp R \mid D, Y$.

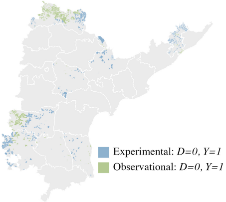
For units on the left, we visualize density of R after PCA on the right.



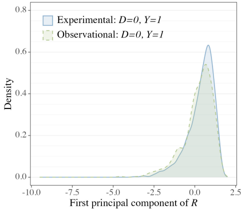
(a) Units with $D = 0$ and $Y = 0$.



(b) Densities of $R \mid S, D = 0, Y = 0$.



(c) Units with $D = 0$ and $Y = 1$.

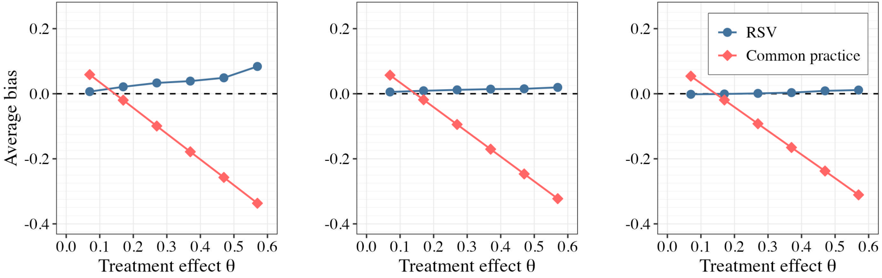


(d) Densities of $R \mid S, D = 0, Y = 1$.

Case study: Synthetic effects, real satellite images

Our method outperforms common practice in terms of average bias.

For each sample size n and synthetic treatment effect value θ,
we use the empirical distribution of $R | Y$.



(a) $n = 1000$ (b) $n = 2000$ (c) $n = 3000$

Case study: Takeaway

substantially reduce bias

Recommendations for practice

How to conduct program evaluation with RSVs?

- auxiliary sample: RSVs with linked outcomes
- three predictions: outcome, treatment, sample
- efficient inference: shortest, prediction-adjusted confidence interval

Which diagnostics should researchers assess?

- weak RSV test
- joint test of identifying assumptions

We would love to talk more!

- email: asheshr@mit.edu; rahul_singh@fas.harvard.edu; dviviano@fas.harvard.edu

Recommendations for practice

How to conduct program evaluation with RSVs?

- auxiliary sample: RSVs with linked outcomes
- three predictions: outcome, treatment, sample
- efficient inference: shortest, prediction-adjusted confidence interval

Which diagnostics should researchers assess?

- weak RSV test
- joint test of identifying assumptions

We would love to talk more!

- email: asheshr@mit.edu; rahul_singh@fas.harvard.edu; dviviano@fas.harvard.edu

Recommendations for practice

How to conduct program evaluation with RSVs?

- auxiliary sample: RSVs with linked outcomes
- three predictions: outcome, treatment, sample
- efficient inference: shortest, prediction-adjusted confidence interval

Which diagnostics should researchers assess?

- weak RSV test
- joint test of identifying assumptions

We would love to talk more!

- email: asheshr@mit.edu; rahul_singh@fas.harvard.edu; dviviano@fas.harvard.edu