Program Evaluation with Remotely Sensed Variables

Ashesh Rambachan
MIT

Rahul Singh
Harvard

Davide Viviano
Harvard

June, 2025

Problem description

- Growing availability of experiments in social science and industry
- But... analysis often challenging due to limited access to outcomes
  - Ex1 Anti-poverty programs require measuring consumption
  - Ex2 Anti-deforestation/land-use programs requiring observing deforestation

Problem description

- Growing availability of experiments in social science and industry
- But... analysis often challenging due to limited access to outcomes

  Ex1 Anti-poverty programs require measuring consumption

  Ex2 Anti-deforestation/land-use programs requiring observing deforestation
- Complement traditional surveys with remotely sensed variables?
  - Nightlights, satellite images, mobile phone data, noisy surveys, etc

Pros Cheap to collect

Cons Noisy and difficult to analyze

Problem description

- Growing availability of experiments in social science and industry
- But... analysis often challenging due to limited access to outcomes

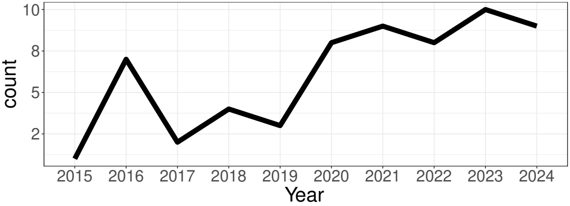
  Ex1 Anti-poverty programs require measuring consumption

  Ex2 Anti-deforestation/land-use programs requiring observing deforestation
- Complement traditional surveys with remotely sensed variables?
  - Nightlights, satellite images, mobile phone data, noisy surveys, etc

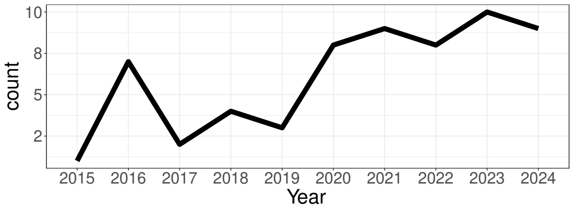
  Pros Cheap to collect

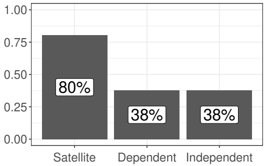
  Cons Noisy and difficult to analyze
- How (and when) should we use such data for program evaluation?

"Top-5" and AEJ Applied econ journals

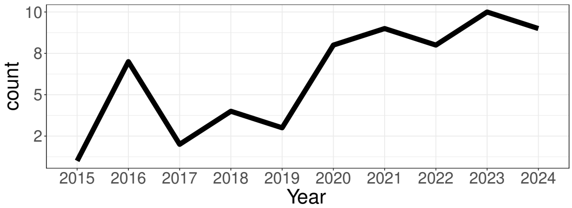


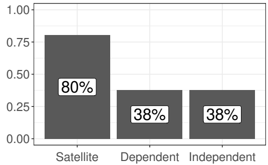
“Top-5” and AEJ Applied econ journals





“Top-5” and AEJ Applied econ journals

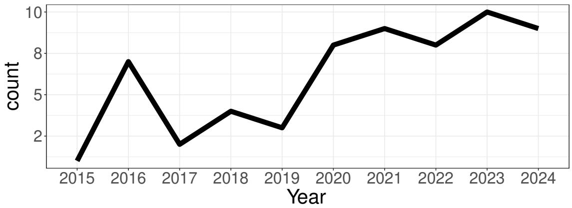


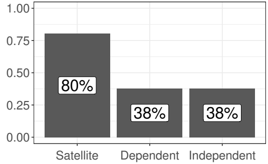


Other examples

- Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) (https://www.devdatalab.org/shrug)

“Top-5” and AEJ Applied econ journals

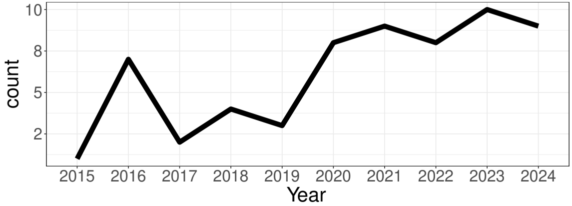


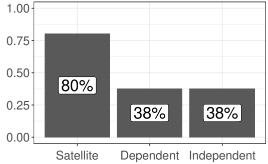


Other examples

- Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) (https://www.devdatalab.org/shrug)
- “Cash for carbon: a randomized trial of payments for ecosystem services to reduce deforestation” (Jayachandran et al., 2017)

"Top-5" and AEJ Applied econ journals





Other examples

- Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) (https://www.devdatalab.org/shrug)
- "Cash for carbon: a randomized trial of payments for ecosystem services to reduce deforestation" (Jayachandran et al., 2017)
- "Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs" (Huang et al., 2021)
- "Estimating Impact with Surveys versus Digital Traces: Evidence fro RCTs in Togo" (Aiken et al, 2023)

This paper in one slide

- Researchers have access to obs study and experiment
  - In the obs study, researchers observe outcomes and RSV
  - In the experiment, researchers observe treatment and RSV

This paper in one slide

- Researchers have access to obs study and experiment
  - In the obs study, researchers observe outcomes and RSV
  - In the experiment, researchers observe treatment and RSV

- In this paper we focus on RSV as a post-outcome variable

  Ex luminosity changes as functions of consumption (and other factors)

This paper in one slide

- Researchers have access to obs study and experiment
  - In the obs study, researchers observe outcomes and RSV
  - In the experiment, researchers observe treatment and RSV

- In this paper we focus on RSV as a post-outcome variable

  Ex luminosity changes as functions of consumption (and other factors)

- Principled way for inference on treatment effects in the experiment

This paper in one slide

- Researchers have access to obs study and experiment
  - In the obs study, researchers observe outcomes and RSV
  - In the experiment, researchers observe treatment and RSV
- In this paper we focus on RSV as a post-outcome variable
  - Ex luminosity changes as functions of consumption (and other factors)
- Principled way for inference on treatment effects in the experiment
  - Comparing predicted outcomes betw/ treated and controls has bias

This paper in one slide

- Researchers have access to obs study and experiment
  - In the obs study, researchers observe outcomes and RSV
  - In the experiment, researchers observe treatment and RSV
- In this paper we focus on RSV as a post-outcome variable

  Ex luminosity changes as functions of consumption (and other factors)
- Principled way for inference on treatment effects in the experiment
  - Comparing predicted outcomes betw/ treated and controls has bias
  - Provide a simple nonparametric identification result for TE
  - Estimation can use arbitrary ML algorithms without affecting inference

This paper in one slide

- Researchers have access to obs study and experiment
  - In the obs study, researchers observe outcomes and RSV
  - In the experiment, researchers observe treatment and RSV
- In this paper we focus on RSV as a post-outcome variable

  Ex luminosity changes as functions of consumption (and other factors)
- Principled way for inference on treatment effects in the experiment
  - Comparing predicted outcomes betw/ treated and controls has bias
  - Provide a simple nonparametric identification result for TE
  - Estimation can use arbitrary ML algorithms without affecting inference
  - Study large-scale public-policy (Smartcards) replicating results in Muralidharan et al., 2023 with satellite images

Related literature in econometrics and statistics

- Growing econometric literature on data fusion
  - Long and short regressions [Cross and Manski, 2002; Molinari and Peski, 2006; D'Haultfoeuille, Gaillac and Maurel, 2024; Bareinboim and Pearl, 2016]
  - Surrogate literature [Athey et al., 2024; Kallus and Mao, 2022] + proxy with surrogates [Imbens et al., 2024; Ghassami et al., 2022]

Related literature in econometrics and statistics

- Growing econometric literature on data fusion
  - Long and short regressions [Cross and Manski, 2002; Molinari and Peski, 2006; D'Haultfoeuille, Gaillac and Maurel, 2024; Bareinboim and Pearl, 2016]
  - Surrogate literature [Athey et al., 2024; Kallus and Mao, 2022] + proxy with surrogates [Imbens et al., 2024; Ghassami et al., 2022]

$\Rightarrow$ None study a missing data problem for the outcome, with auxiliary dataset with a post-outcome variable

Related literature in econometrics and statistics

- Growing econometric literature on data fusion
  - Long and short regressions [Cross and Manski, 2002; Molinari and Peski, 2006; D'Haultfoeuille, Gaillac and Maurel, 2024; Bareinboim and Pearl, 2016]
  - Surrogate literature [Athey et al., 2024; Kallus and Mao, 2022] + proxy with surrogates [Imbens et al., 2024; Ghassami et al., 2022]

  ⇒ None study a missing data problem for the outcome, with auxiliary dataset with a post-outcome variable

- Generative models [Gentzkow, Shapiro and Taddy, 2019; Battaglia et al., 2024]

Related literature in econometrics and statistics

- Growing econometric literature on data fusion
  - Long and short regressions [Cross and Manski, 2002; Molinari and Peski, 2006; D'Haultfoeuille, Gaillac and Maurel, 2024; Bareinboim and Pearl, 2016]
  - Surrogate literature [Athey et al., 2024; Kallus and Mao, 2022] + proxy with surrogates [Imbens et al., 2024; Ghassami et al., 2022]

  ⇒ None study a missing data problem for the outcome, with auxiliary dataset with a post-outcome variable

- Generative models [Gentzkow, Shapiro and Taddy, 2019; Battaglia et al., 2024]

  ⇒ here no need of correctly specified model for RSV

Related literature in econometrics and statistics

- Growing econometric literature on data fusion
  - Long and short regressions [Cross and Manski, 2002; Molinari and Peski, 2006; D'Haultfoeuille, Gaillac and Maurel, 2024; Bareinboim and Pearl, 2016]
  - Surrogate literature [Athey et al., 2024; Kallus and Mao, 2022] + proxy with surrogates [Imbens et al., 2024; Ghassami et al., 2022]

  ⇒ None study a missing data problem for the outcome, with auxiliary dataset with a post-outcome variable

- Generative models [Gentzkow, Shapiro and Taddy, 2019; Battaglia et al., 2024]

  ⇒ here no need of correctly specified model for RSV

- Measurement error and ML imputations
  - Classic ME [Hasuman, 2001; Hu, 2008; Molinari, 2008; Schennach, 2020; ...]
  - Outcome or covariates imputation [Alliott et al., 2023; Angelopoulos et al., (2023); Egami et al. (2024); Zhang et al. (2023); ...]

Related literature in econometrics and statistics

- Growing econometric literature on data fusion
  - Long and short regressions [Cross and Manski, 2002; Molinari and Peski, 2006; D'Haultfoeuille, Gaillac and Maurel, 2024; Bareinboim and Pearl, 2016]
  - Surrogate literature [Athey et al., 2024; Kallus and Mao, 2022] + proxy with surrogates [Imbens et al., 2024; Ghassami et al., 2022]

  ⇒ None study a missing data problem for the outcome, with auxiliary dataset with a post-outcome variable

- Generative models [Gentzkow, Shapiro and Taddy, 2019; Battaglia et al., 2024]

  ⇒ here no need of correctly specified model for RSV

- Measurement error and ML imputations
  - Classic ME [Hasuman, 2001; Hu, 2008; Molinari, 2008; Schennach, 2020; ...]
  - Outcome or covariates imputation [Alliott et al., 2023; Angelopoulos et al., (2023); Egami et al. (2024); Zhang et al. (2023); ...]

  ⇒ Here ME problem with missing data (data fusion)

Content

Notation

- $D \in \{0, 1\}$: binary treatment
- $Y(1), Y(0)$ POs under treatment and control satisfy SUTVA
- $S \in \{e, o\}$: observation is in experiment or obs study
- R is the remote sensed variable (e.g., satellite images)

Notation

- $D \in \{0, 1\}$: binary treatment
- $Y(1), Y(0)$ POs under treatment and control satisfy SUTVA
- $S \in \{e, o\}$: observation is in experiment or obs study
- R is the remote sensed variable (e.g., satellite images)

- Researchers observe n independent observations

$$\Big(\underbrace{Y_i 1\{S_i = o\}}_{\text{only on obs study}}, D_i, R_i, S_i \Big)_{i=1}^{n}$$

Notation

- $D \in \{0, 1\}$: binary treatment
- $Y(1), Y(0)$ POs under treatment and control satisfy SUTVA
- $S \in \{e, o\}$: observation is in experiment or obs study
- R is the remote sensed variable (e.g., satellite images)

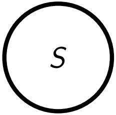
- Researchers observe n independent observations

$$\Big(\underbrace{Y_i 1\{S_i = o\}}_{\text{only on obs study}}, D_i, R_i, S_i \Big)_{i=1}^{n}$$

- Some extensions
  - Pre-treatment covariates X
  - Researchers observe *some* outcomes in the experiment

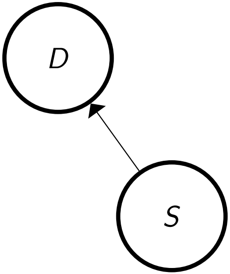
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



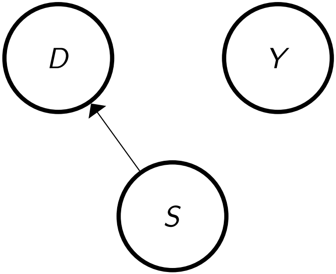
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



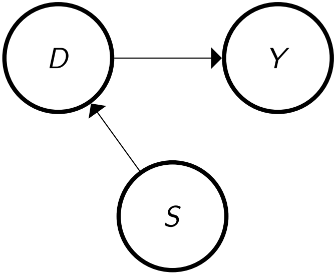
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



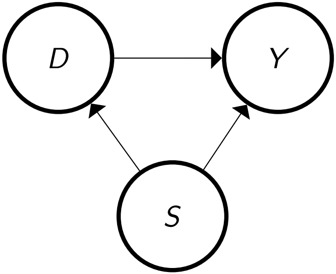
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



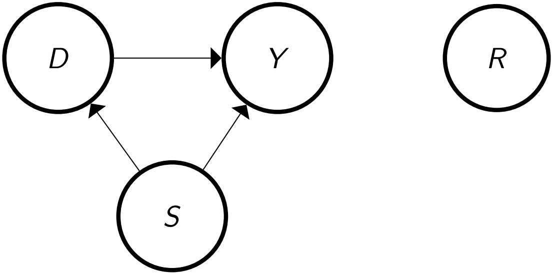
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



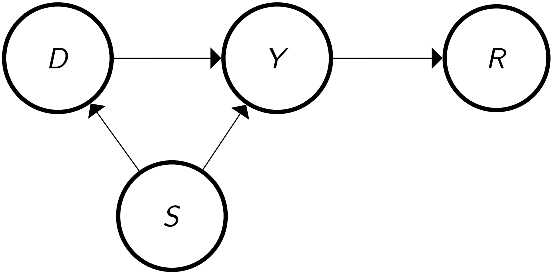
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



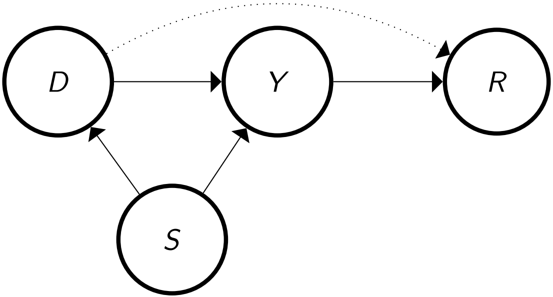
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



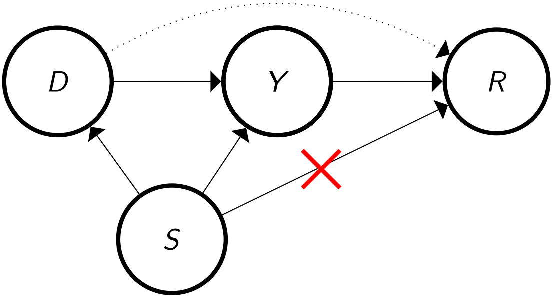
Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



Causal model for RSV

- Goal: $\theta = \mathbb{E}[Y(1) - Y(0)|S = E]$ (ATE in experiment)



Preliminary conditions on experiment and obs/ study

Condition 1: randomization occurs in the experiment

- $D \perp \Big(Y(1), Y(0)\Big) | S = e$ and $P(D = 1 | S = e) \in (0, 1)$

Preliminary conditions on experiment and obs/ study

Condition 1: randomization occurs in the experiment

- $D \perp \Big(Y(1), Y(0)\Big) | S = e$ and $P(D = 1 | S = e) \in (0, 1)$

Condition 2: Obs study

- Either
  - Some units are treated in obs study: $P(D = 1 | S = o) \in (0, 1)$
  - or treatment has no direct effect on RSV, $D \perp R | Y$.

Preliminary conditions on experiment and obs/ study

Condition 1: randomization occurs in the experiment

- $D \perp \Big(Y(1), Y(0)\Big) | S = e$ and $P(D = 1 | S = e) \in (0, 1)$

Condition 2: Obs study

- Either
  - Some units are treated in obs study: $P(D = 1 | S = o) \in (0, 1)$
  - or treatment has no direct effect on RSV, $D \perp R | Y$.

Condition 3: Stability $S \perp R | Y, D$

Remarks and desiderata

(1) Direct effect and stability are jointly testable

(2) If you expect direct effects:

- Collect some outcomes in the treatment group
- Use a larger vector of outcomes in your application

Remarks and desiderata

(1) Direct effect and stability are jointly testable

(2) If you expect direct effects:
- Collect some outcomes in the treatment group
- Use a larger vector of outcomes in your application

Desiderata:

(1) Estimator with infinite variance if $R \perp Y$ (cannot learn much)

(2) Consistent estimator if $R \not\perp Y$

(3) Do not require to know or specify $Y \to R$

Remarks and desiderata

(1) Direct effect and stability are jointly testable
(2) If you expect direct effects:
- Collect some outcomes in the treatment group
- Use a larger vector of outcomes in your application

Desiderata:

(1) Estimator with infinite variance if $R \perp Y$ (cannot learn much)
(2) Consistent estimator if $R \not\perp Y$
(3) Do not require to know or specify $Y \to R$
! Bonus: efficient choice of R

Content

Surrogate method: what we may be tempted to do

- Intuitive estimator would be:

Surrogate method: what we may be tempted to do

- Intuitive estimator would be:
  - Estimate $f(R) = \mathbb{E}[Y|R, S = o]$ in the observational study (assuming here no observational unit exposed to treatment)

Surrogate method: what we may be tempted to do

- Intuitive estimator would be:
  - Estimate $f(R) = \mathbb{E}[Y|R, S = o]$ in the observational study (assuming here no observational unit exposed to treatment)
  - Estimate ATE by taking treated/control differences relative to the prediction $f(R)$ in the experiment:

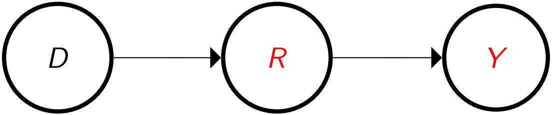
$$\tilde{\theta} = \mathbb{E}[f(R)|D = 1, S = E] - \mathbb{E}[f(R)|D = 0, S = E]$$

Surrogate method: what we may be tempted to do

- Intuitive estimator would be:
  - Estimate $f(R) = \mathbb{E}[Y|R, S = o]$ in the observational study (assuming here no observational unit exposed to treatment)
  - Estimate ATE by taking treated/control differences relative to the prediction $f(R)$ in the experiment:

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

  - Exercise mimics surrogates in data fusion lit (Athey et al., 2018)



Surrogate method is biased with RSV

- Counter-argument

$$R = Y\beta + \varepsilon, \quad Y = \theta D + \eta,$$

Surrogate method is biased with RSV

- Counter-argument

$$R = Y\beta + \varepsilon, \quad Y = \theta D + \eta, \quad f(R) = R\beta^*, \quad \beta^* = \frac{\mathrm{Cov}(Y, R)}{\mathrm{Var}(R)}$$

Surrogate method is biased with RSV

- Counter-argument

$$R = Y\beta + \varepsilon, \quad Y = \theta D + \eta, \quad f(R) = R\beta^*, \quad \beta^* = \frac{\mathrm{Cov}(Y, R)}{\mathrm{Var}(R)}$$

Then with some algebra

$$\widetilde{\theta} = \beta^* \Big(\mathbb{E}[R|D = 1, S = e] - \mathbb{E}[R|D = 0, S = e] \Big)$$

Surrogate method is biased with RSV

- Counter-argument

$$R = Y\beta + \varepsilon, \quad Y = \theta D + \eta, \quad f(R) = R\beta^*, \quad \beta^* = \frac{\mathrm{Cov}(Y, R)}{\mathrm{Var}(R)}$$

Then with some algebra

$$\widetilde{\theta} = \beta^* \Big(\mathbb{E}[R|D = 1, S = e] - \mathbb{E}[R|D = 0, S = e] \Big) = \beta^* \beta \theta$$

$\beta^* \beta = \frac{\mathrm{Cov}(Y,R)^2}{\mathrm{Var}(Y)\mathrm{Var}(R)}$ is squared correlation coefficient

Surrogate method is biased with RSV

- Counter-argument

$$R = Y\beta + \varepsilon, \quad Y = \theta D + \eta, \quad f(R) = R\beta^*, \quad \beta^* = \frac{\mathrm{Cov}(Y, R)}{\mathrm{Var}(R)}$$

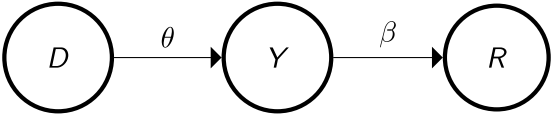
Then with some algebra

$$\widetilde{\theta} = \beta^* \Big(\mathbb{E}[R|D = 1, S = e] - \mathbb{E}[R|D = 0, S = e] \Big) = \beta^* \beta \theta$$

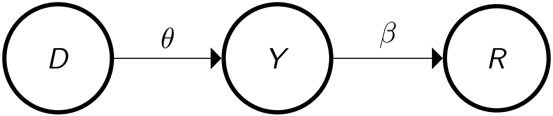
$\beta^* \beta = \frac{\mathrm{Cov}(Y,R)^2}{\mathrm{Var}(Y)\mathrm{Var}(R)}$ is squared correlation coefficient

$\Rightarrow$ Bias can be significant in applications (E.g., in a cash transfer program to reduce crop burning by farmers, estimated effects using satellite images under-estimate ATE by at least 50%)

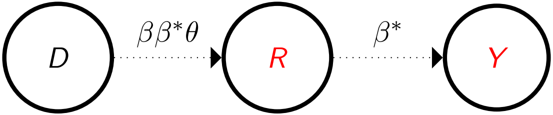
Surrogate method: intuition



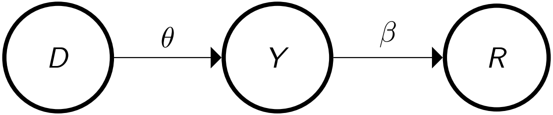
Surrogate method: intuition



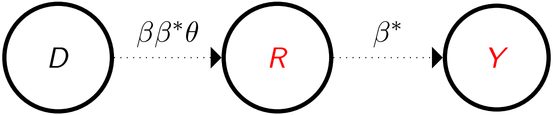
Surrogate:



Surrogate method: intuition

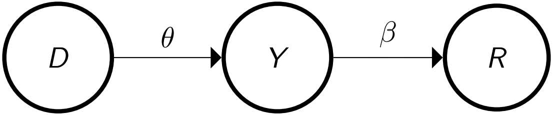


Surrogate:

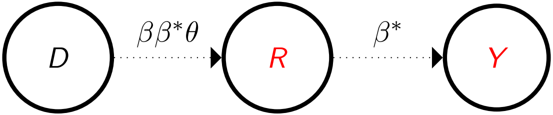


$\Rightarrow$ If $\beta = 0$, estimated effect is a precise zero

Surrogate method: intuition



Surrogate:



⇒ If $\beta = 0$, estimated effect is a precise zero

Prop There exist at least two DGPs satisfying $(D, S) \perp R | Y$ and Conditions 1-3 hold, so that (formula)

$$\underbrace{\theta}_{\text{estimand}} - \underbrace{\tilde{\theta}}_{\text{surrogate prediction}}$$

can have arbitrary sign reversals under RSV assumptions.

Content

Constructive identification argument

- Suppose $Y \in \{0, 1\}$ and no direct effects $D \perp R | Y$.

Constructive identification argument

- Suppose $Y \in \{0, 1\}$ and no direct effects $D \perp R | Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R = r | D = d, S = e)}_{\text{we can identify}}$$

Constructive identification argument

- Suppose $Y \in \{0,1\}$ and no direct effects $D \perp R|Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R=r|D=d,S=e)}_{\text{we can identify}}$$

$$= \sum_{y\in\{0,1\}}$$

Constructive identification argument

- Suppose $Y \in \{0,1\}$ and no direct effects $D \perp R | Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R = r | D = d, S = e)}_{\text{we can identify}}$$

$$= \sum_{y \in \{0,1\}} \underbrace{P(R = r | Y = y, D = d, S = e)}_{\text{not identified}}$$

Constructive identification argument

- Suppose $Y \in \{0,1\}$ and no direct effects $D \perp R | Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R = r | D = d, S = e)}_{\text{we can identify}}$$

$$= \sum_{y \in \{0,1\}} \underbrace{P(R = r | Y = y, D = d, S = e)}_{\text{not identified}} \underbrace{P(Y(d) = y | S = e)}_{\text{estimand}}$$

Constructive identification argument

- Suppose $Y \in \{0,1\}$ and no direct effects $D \perp R | Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R = r | D = d, S = e)}_{\text{we can identify}}$$

$$= \sum_{y \in \{0,1\}} \underbrace{P(R = r | Y = y, D = d, S = e)}_{\text{not identified}} \underbrace{P(Y(d) = y | S = e)}_{\text{estimand}}$$

  - Now use the stability assumption (+ no direct effects)

$$P(R = r | Y = y, D = d, \mathbf{S} = \mathbf{e})$$

Constructive identification argument

- Suppose $Y \in \{0,1\}$ and no direct effects $D \perp R | Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R = r | D = d, S = e)}_{\text{we can identify}}$$

$$= \sum_{y \in \{0,1\}} \underbrace{P(R = r | Y = y, D = d, S = e)}_{\text{not identified}} \underbrace{P(Y(d) = y | S = e)}_{\text{estimand}}$$

  - Now use the stability assumption (+ no direct effects)

$$P(R = r | Y = y, D = d, \mathbf{S} = \mathbf{e}) = P(R = r | Y = y, \mathbf{S} = \mathbf{o})$$

Constructive identification argument

- Suppose $Y \in \{0,1\}$ and no direct effects $D \perp R | Y$.
- Step one is to recognize that we identify a mixture:
  - The probability of the image depends on the prob of each PO:

$$\underbrace{P(R=r|D=d,S=e)}_{\text{we can identify}}$$
$$= \sum_{y \in \{0,1\}} \underbrace{P(R=r|Y=y,D=d,S=e)}_{\text{not identified}} \underbrace{P(Y(d)=y|S=e)}_{\text{estimand}}$$

  - Now use the stability assumption (+ no direct effects)

$$P(R=r|Y=y,D=d,\mathbf{S}=\mathbf{e}) = P(R=r|Y=y,\mathbf{S}=\mathbf{o})$$

$\Rightarrow$ The probability of what we observe in experiment is a mixture of probabilities of what we would have observed in the obs study

Identification: second step

- After taking differences, we can write

$$\underbrace{P(R = r|D = 1, S = e) - P(R = r|D = 0, S = e)}_{\text{How RSV changes if we change treatment}}$$

Identification: second step

- After taking differences, we can write

$$\underbrace{P(R=r|D=1,S=e)-P(R=r|D=0,S=e)}_{\text{How RSV changes if we change treatment}}$$

$$=\boldsymbol{\theta}$$

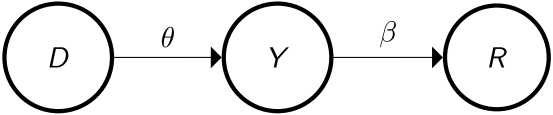
Identification: second step

- After taking differences, we can write

$$\underbrace{P(R=r|D=1,S=e)-P(R=r|D=0,S=e)}_{\text{How RSV changes if we change treatment}}$$
$$=\boldsymbol{\theta}\underbrace{\Big(P(R=r|Y=1,S=o)-P(R=r|Y=0,S=o)\Big)}_{\text{How RSV changes if we change outcome:}\beta}.$$

- After taking differences, we can write

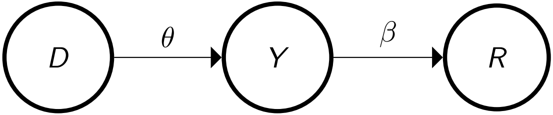
$$\underbrace{P(R=r|D=1,S=e)-P(R=r|D=0,S=e)}_{\text{How RSV changes if we change treatment}}$$
$$=\boldsymbol{\theta}\underbrace{\Big(P(R=r|Y=1,S=o)-P(R=r|Y=0,S=o)\Big)}_{\text{How RSV changes if we change outcome:}\beta}.$$



Identification: second step

- After taking differences, we can write

$$\underbrace{P(R=r|D=1,S=e)-P(R=r|D=0,S=e)}_{\text{How RSV changes if we change treatment}}$$
$$=\boldsymbol{\theta}\underbrace{\Big(P(R=r|Y=1,S=o)-P(R=r|Y=0,S=o)\Big)}_{\text{How RSV changes if we change outcome:}\beta}.$$



$\Rightarrow$ Challenge: such probabilities are difficult to predict.

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\left[\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}\right]}_{\text{How RSV changes if we change treatment}}$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\left[\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}\right]}_{\text{How RSV changes if we change treatment}} \times P(R)$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\left[\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}\right]}_{\text{How RSV changes if we change treatment}} \times P(R)$$

$$= \boldsymbol{\theta}$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\left[\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}\right]}_{\text{How RSV changes if we change treatment}} \times P(R)$$

$$= \boldsymbol{\theta} \underbrace{\left(\frac{P(Y=1,S=o|R)}{(Y=1,S=o)} - \frac{P(Y=0,S=o|R)}{P(Y=0,S=o)}\right)}_{\text{How RSV changes if we change outcome}} \times$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes rule. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\left[\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}\right]}_{\text{How RSV changes if we change treatment}} \times P(R)$$

$$= \boldsymbol{\theta} \underbrace{\left(\frac{P(Y=1,S=o|R)}{(Y=1,S=o)} - \frac{P(Y=0,S=o|R)}{P(Y=0,S=o)}\right)}_{\text{How RSV changes if we change outcome}} \times P(R).$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes theorem. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}}_{\text{How RSV changes if we change treatment}} \times \cancel{P(R)}$$

$$= \boldsymbol{\theta} \underbrace{\left(\frac{P(Y=1,S=o|R)}{(Y=1,S=o)} - \frac{P(Y=0,S=o|R)}{P(Y=0,S=o)} \right)}_{\text{How RSV changes if we change outcome}} \times \cancel{P(R)}.$$

Identification: obtaining simple moment conditions

- The following step is to use Bayes theorem. For instance

$$P(R|D=1,S=e) = \frac{P(D=1,S=e|R)}{P(D=1,S=e)} \times \underbrace{P(R)}_{\text{hard to estimate}}$$

- Good news is that $P(R)$ is on both sides

$$\underbrace{\frac{P(D=1,S=e|R)}{P(D=1,S=e)} - \frac{P(D=0,S=e|R)}{P(D=0,S=e)}}_{\text{How RSV changes if we change treatment}} \times \cancel{P(R)}$$

$$= \boldsymbol{\theta} \underbrace{\left(\frac{P(Y=1,S=o|R)}{(Y=1,S=o)} - \frac{P(Y=0,S=o|R)}{P(Y=0,S=o)} \right)}_{\text{How RSV changes if we change outcome}} \times \cancel{P(R)}.$$

Conditional moments that do not depend on $P(R)$

Can we circunvent correct specification of $D|R$, $Y|R$?

Identification: conditional moments

- Write
$$\Delta_i(e) = \frac{1\{D=1, S=e\}}{P(D=1, S=e)} - \frac{1\{D=0, S=e\}}{P(D=0, S=e)}$$
and similarly $\Delta_i(o) = \frac{1\{Y=1,S=o\}}{P(Y=1,S=o)} - \frac{1\{Y=0,S=o\}}{P(Y=0,S=o)}$

Identification: conditional moments

- Write

$$\Delta_i(e) = \frac{1\{D=1, S=e\}}{P(D=1, S=e)} - \frac{1\{D=0, S=e\}}{P(D=0, S=e)}$$

and similarly $\Delta_i(o) = \frac{1\{Y=1,S=o\}}{P(Y=1,S=o)} - \frac{1\{Y=0,S=o\}}{P(Y=0,S=o)}$

Thm (Informal) Identification simplifies to

$$\mathbb{E}[\Delta_i(e) - \theta\Delta_i(o)|R] = 0$$

$\Rightarrow$ No need to know data generating process for $R|Y$

A simple Wald estimand

- Take any representation of R, call it $H(R)$ [e.g. average luminosity]
- Require that $\mathbb{E}[H(R_i)\Delta_i(o)] \neq 0 \Rightarrow$ predictive of the outcome

A simple Wald estimand

- Take any representation of R, call it $H(R)$ [e.g. average luminosity]
- Require that $\mathbb{E}[H(R_i)\Delta_i(o)] \neq 0 \Rightarrow$ predictive of the outcome

- Then we can write

$$\theta = \underbrace{\mathbb{E}[H(R_i)\Delta_i(e)]}_{\text{Treatment change predicted by RSV}} \times \underbrace{\mathbb{E}[H(R_i)\Delta_i(o)]^{-1}}_{\text{Outcome change predicted by RSV}}$$

A simple Wald estimand

- Take any representation of R, call it $H(R)$ [e.g. average luminosity]
- Require that $\mathbb{E}[H(R_i)\Delta_i(o)] \neq 0 \Rightarrow$ predictive of the outcome

- Then we can write

$$\theta = \underbrace{\mathbb{E}[H(R_i)\Delta_i(e)]}_{\text{Treatment change predicted by RSV}} \times \underbrace{\mathbb{E}[H(R_i)\Delta_i(o)]^{-1}}_{\text{Outcome change predicted by RSV}}$$

- Intuition:
  - See how much R and D commove
  - Note that some of this is attributed to effect of Y to R
  - Appropriately divide by how much Y and R commove

A simple Wald estimand

- Take any representation of R, call it $H(R)$ [e.g. average luminosity]
- Require that $\mathbb{E}[H(R_i)\Delta_i(o)] \neq 0 \Rightarrow$ predictive of the outcome
- Then we can write

$$\theta = \underbrace{\mathbb{E}[H(R_i)\Delta_i(e)]}_{\text{Treatment change predicted by RSV}} \times \underbrace{\mathbb{E}[H(R_i)\Delta_i(o)]^{-1}}_{\text{Outcome change predicted by RSV}}$$

- Intuition:
  - See how much R and D commove
  - Note that some of this is attributed to effect of Y to R
  - Appropriately divide by how much Y and R commove
- Testable implication:
  - θ is constant as we vary $H(\cdot)$
  - Can change representation $H(\cdot)$ to test if estimate changes

Content

Efficient representation

- How to choose $H(R)$ to maximize precision?
  - Average luminosity, weighted luminosity, prediction of $Y|R$?

Efficient representation

- How to choose $H(R)$ to maximize precision?
  - Average luminosity, weighted luminosity, prediction of $Y|R$?
- Start from the conditional moment as

$$\Delta_i(e) = \theta \Delta_i(o) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i | R_i] = 0$$

Efficient representation

- How to choose $H(R)$ to maximize precision?
  - Average luminosity, weighted luminosity, prediction of $Y|R$?
- Start from the conditional moment as

$$\Delta_i(e) = \theta \Delta_i(o) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i | R_i] = 0$$

- Representation as an optimal instrument [Chamberlain, 1987; Newey, 1993]

$$H(R) = \frac{\mathbb{E}[\Delta_i(o)|R]}{\sigma^2(R)}, \quad \sigma^2(R) = \mathbb{E}[\varepsilon^2|R].$$

Efficient representation

- How to choose $H(R)$ to maximize precision?
  - Average luminosity, weighted luminosity, prediction of $Y|R$?
- Start from the conditional moment as

$$\Delta_i(e) = \theta\Delta_i(o) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i|R_i] = 0$$

- Representation as an optimal instrument [Chamberlain, 1987; Newey, 1993]

$$H(R) = \frac{\mathbb{E}[\Delta_i(o)|R]}{\sigma^2(R)}, \quad \sigma^2(R) = \mathbb{E}[\varepsilon^2|R].$$

- Intuition:
  - use a function of the predicted outcome as an instrument
  - Heteroskedasticity depending on treatment and outcome predictions

Algorithm in a nutschell

- Use a high-dim estimator with regularization to estimate

$$\underbrace{P(Y=1|R,S=o)}_{\text{outcome prediction}},\quad \underbrace{P(D=1|R,S=e)}_{\text{treatment prediction}},\quad \underbrace{P(S=e|R)}_{\text{assignment prediction}}$$

- Use these estimates to estimate $\hat{H}(R)$

Algorithm in a nutschell

- Use a high-dim estimator with regularization to estimate

$$\underbrace{P(Y=1|R,S=o)}_{\text{outcome prediction}},\quad \underbrace{P(D=1|R,S=e)}_{\text{treatment prediction}},\quad \underbrace{P(S=e|R)}_{\text{assignment prediction}}$$

- Use these estimates to estimate $\hat{H}(R)$
- Obtain your estimated effect

$$\hat{\theta}=\frac{\mathbb{E}_n[\hat{\Delta}_i(e)\hat{H}(R_i)]}{\mathbb{E}_n[\hat{\Delta}_i(o)\hat{H}(R_i)]}$$

Algorithm in a nutschell

- Use a high-dim estimator with regularization to estimate

$$\underbrace{P(Y=1|R,S=o)}_{\text{outcome prediction}},\quad \underbrace{P(D=1|R,S=e)}_{\text{treatment prediction}},\quad \underbrace{P(S=e|R)}_{\text{assignment prediction}}$$

- Use these estimates to estimate $\hat{H}(R)$
- Obtain your estimated effect

$$\hat{\theta} = \frac{\mathbb{E}_n[\hat{\Delta}_i(e)\hat{H}(R_i)]}{\mathbb{E}_n[\hat{\Delta}_i(o)\hat{H}(R_i)]}$$

$\Rightarrow$ Methods of moments with estimated "optimal instrument" [e.g., Newey, 1993]:

- standard $n^{-1/2}$-inference if $\hat{H} \to H^*$ for some arbitrary H^*, with $\mathbb{E}[\Delta_i(o)H^*(R_i)] \neq 0$ and regularity conditions or cross-fitting

Steps for practice

Step 1: choose your target

- Choose outcomes under stability assumption (binary/discrete)

Steps for practice

Step 1: choose your target

- Choose outcomes under stability assumption (binary/discrete)

Step 2: training:

- Train the predictive model to predict outcomes and treatments
- Use these predictions to form optimal representation $H(R)$

Steps for practice

Step 1: choose your target

- Choose outcomes under stability assumption (binary/discrete)

Step 2: training:

- Train the predictive model to predict outcomes and treatments
- Use these predictions to form optimal representation $H(R)$

Step 3: warning signs

- The correlation between $H(R)$ and $\Delta_i(o)$ is "weak"
- Results should remain robust once we change your representation
- If we expect direct effects, need some outcomes exposed to treatment

Steps for practice

Step 1: choose your target

- Choose outcomes under stability assumption (binary/discrete)

Step 2: training:

- Train the predictive model to predict outcomes and treatments
- Use these predictions to form optimal representation $H(R)$

Step 3: warning signs

- The correlation between $H(R)$ and $\Delta_i(o)$ is “weak”
- Results should remain robust once we change your representation
- If we expect direct effects, need some outcomes exposed to treatment

Step 4: estimation and inference as for standard methods of moments

Content

Experimental background

- Can we use remote sensed data to evaluate anti-poverty programs?
- Focus here on the Smartcards experiment [Muralidharan et al., 2016; 2023]
  - How to deliver payments securely to targeted beneficiaries?
  - Smartcards program provided biometrically authenticated payments
  - Large-scale evaluation in Indian state of Andhra Pradesh (2010-2012)

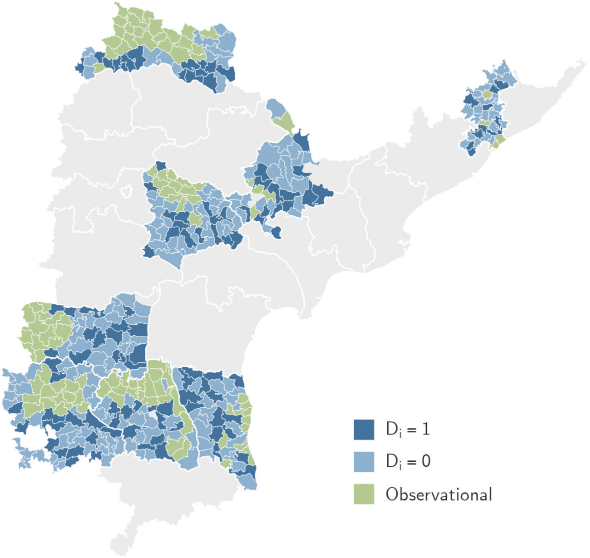
Experimental background

- Can we use remote sensed data to evaluate anti-poverty programs?
- Focus here on the Smartcards experiment [Muralidharan et al., 2016; 2023]
  - How to deliver payments securely to targeted beneficiaries?
  - Smartcards program provided biometrically authenticated payments
  - Large-scale evaluation in Indian state of Andhra Pradesh (2010-2012)
- Subdistricts (mandals) assigned to (more)
  - Not in the study/not randomized (105 mandals)
  - Control – either buffer (136) or pure control (44)
  - Treated – smartcards implemented in in 2010 (111)

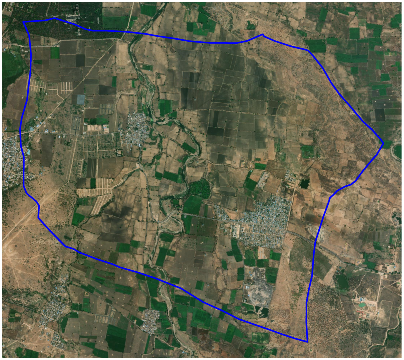
Experimental background

- Can we use remote sensed data to evaluate anti-poverty programs?
- Focus here on the Smartcards experiment [Muralidharan et al., 2016; 2023]
  - How to deliver payments securely to targeted beneficiaries?
  - Smartcards program provided biometrically authenticated payments
  - Large-scale evaluation in Indian state of Andhra Pradesh (2010-2012)
- Subdistricts (mandals) assigned to (more)
  - Not in the study/not randomized (105 mandals)
  - Control – either buffer (136) or pure control (44)
  - Treated – smartcards implemented in in 2010 (111)
- Use village-level information and satellite images to estimate ITT
  - ⇒ We compare point estimates using limited outcome info, with regression that have access to outcome data from the experiment

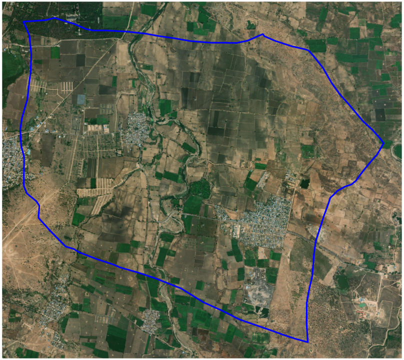
Map from the original experiment

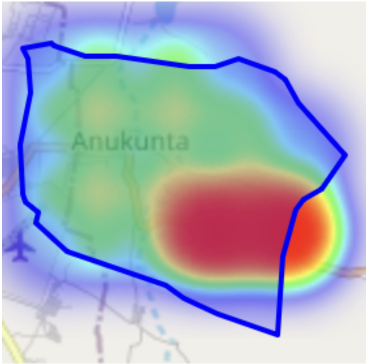


Feature extraction (MOSAIKS features)



Feature extraction (MOSAIKS features)





Validation exercise

Metrics:

- Whether consumption is below its first quartile
- whether no individual in the village has income above Rs. 5,000
- whether no individual in the village has income above Rs 10,000

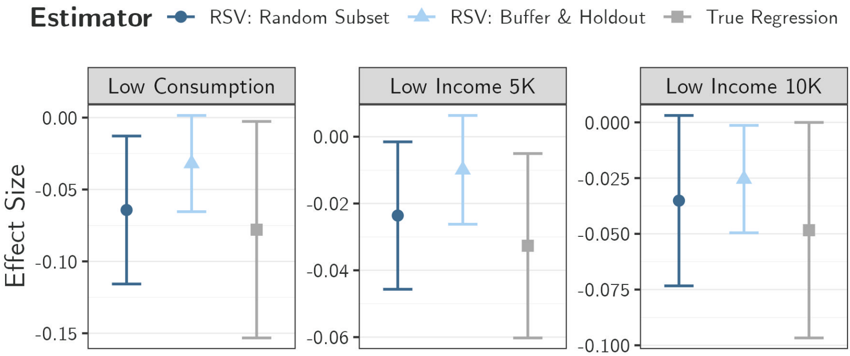
⇒ We can observe outcome metrics for each village using census data

⇒ Compare to regression that uses outcomes of all villages

Two validation exercise implementing our method that

- Uses outcomes for **first half** of the experimental mandals **only**
- Uses outcomes of **buffer** control and **holdout** sample **only**
  - ⇒ Not require outcome info for $\sim$ 3,000 villages

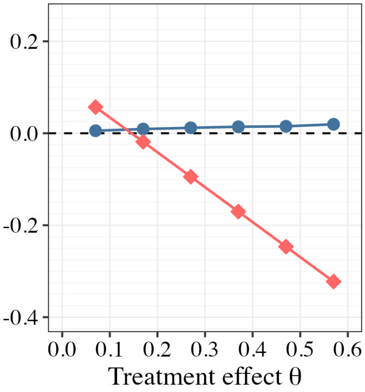
Results



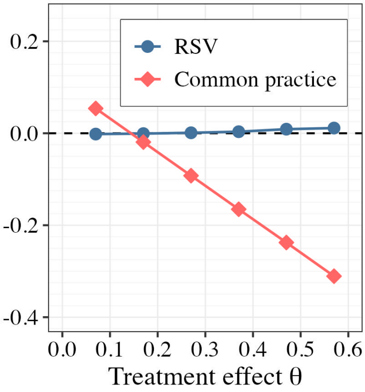
Estimator
RSV: Random Subset
RSV: Buffer & Holdout
True Regression
Low Consumption
Low Income 5K
Low Income 10K
Effect Size
0.00
-0.05
-0.10
-0.15
0.00
-0.02
-0.04
-0.06
0.000
-0.025
-0.050
-0.075
-0.100

Simulations using satellites and consumption data in India

(a) Bias ($n = 2000$)

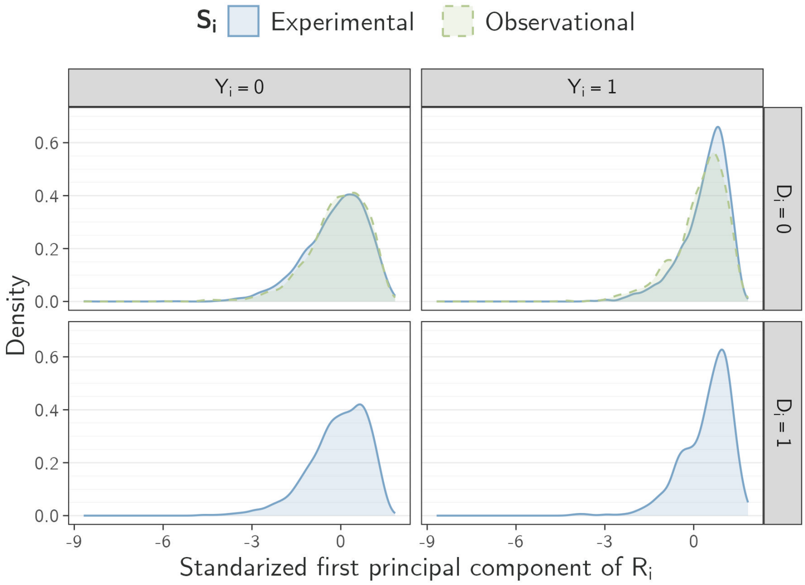


(b) Bias ($n = 3000$)



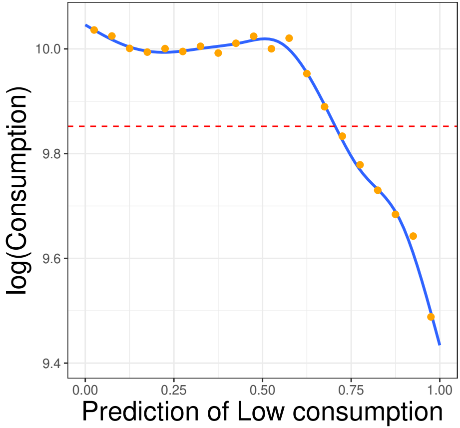
(similar behavior for mean-squared error) (more)

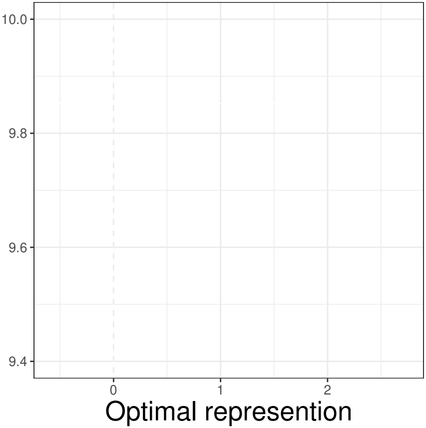
Checking for our assumptions using the full data



Binary outcome prediction using ML

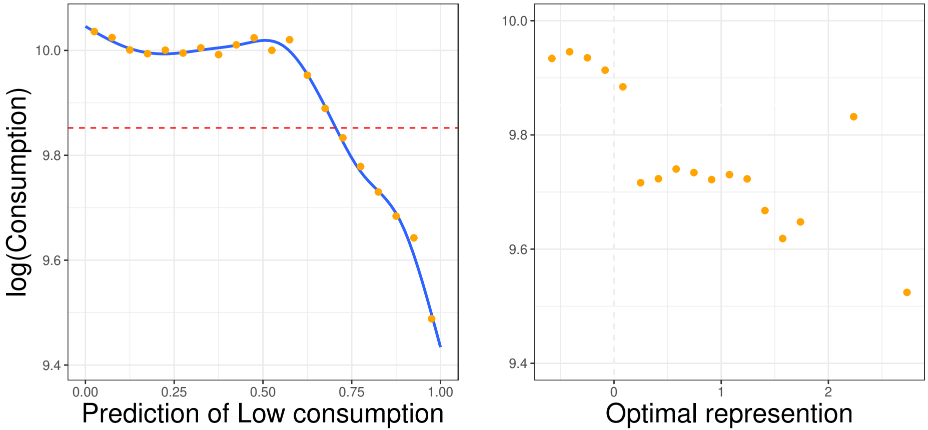
⇒ Per-capita average consumption below 1st quartile





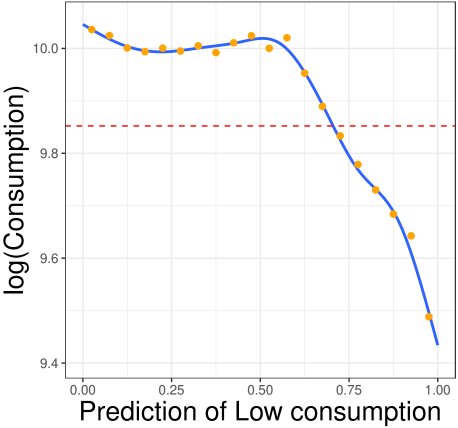
Binary outcome prediction using ML

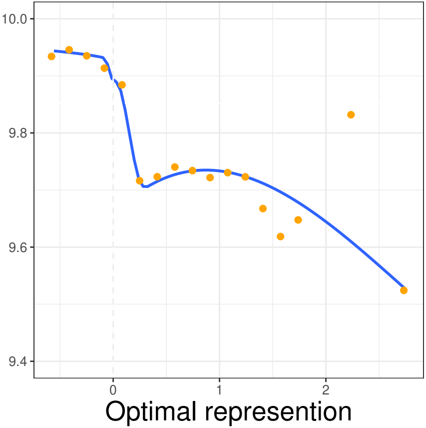
⇒ Per-capita average consumption below 1st quartile



Binary outcome prediction using ML

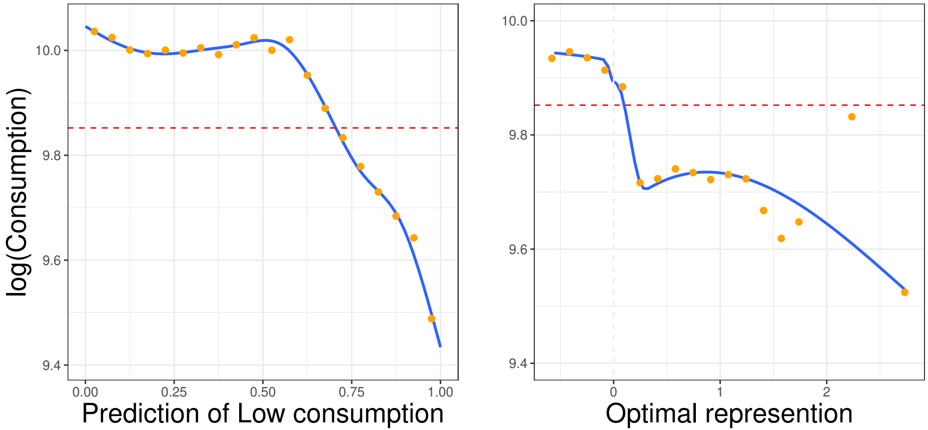
⇒ Per-capita average consumption below 1st quartile





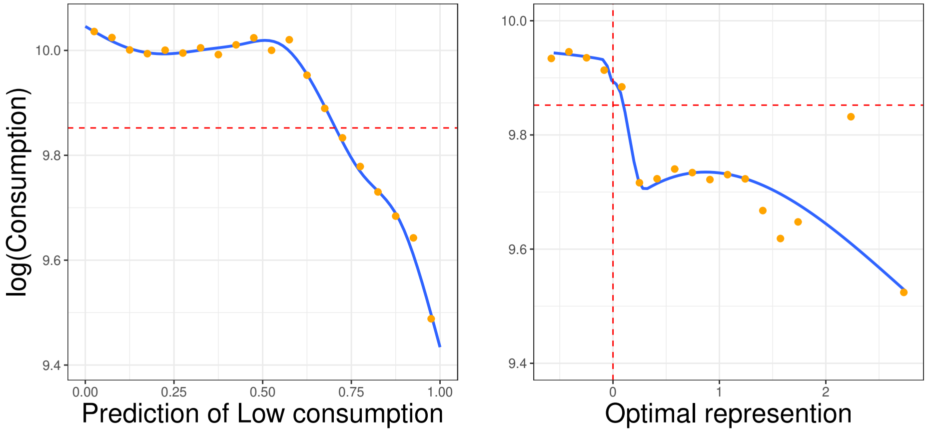
Binary outcome prediction using ML

⇒ Per-capita average consumption below 1st quartile



Binary outcome prediction using ML

⇒ Per-capita average consumption below 1st quartile



Content

Some extensions

Non-binary outcomes:

- assuming no direct effects for simplicity

$$\mathbb{E}\left[\left.\frac{1\{D=d,S=e\}}{P(D=d,S=e)}-\sum_{y\in\mathcal{Y}}\underbrace{\frac{1\{Y=y,S=o\}}{P(Y=y,S=o)}}_{\text{weight: } W_y(o)}\underbrace{P(Y(d)=y|S=e)}_{\text{estimand}}\right|R\right]=0$$

Some extensions

Non-binary outcomes:

- assuming no direct effects for simplicity

$$\mathbb{E}\left[\frac{1\{D=d,S=e\}}{P(D=d,S=e)} - \sum_{y\in\mathcal{Y}} \underbrace{\frac{1\{Y=y,S=o\}}{P(Y=y,S=o)}}_{\text{weight: } W_y(o)} \underbrace{P(Y(d)=y|S=e)}_{\text{estimand}} \Bigg| R\right] = 0$$

$\Rightarrow$ $H(R)$ of dimension $|\mathcal{Y}|$ where $\mathbb{E}[H(R)W(o)]$ is full rank (invertible)

Some extensions

Non-binary outcomes:

- assuming no direct effects for simplicity

$$\mathbb{E}\left[\frac{1\{D=d,S=e\}}{P(D=d,S=e)} - \sum_{y\in\mathcal{Y}} \underbrace{\frac{1\{Y=y,S=o\}}{P(Y=y,S=o)}}_{\text{weight: } W_y(o)} \underbrace{P(Y(d)=y|S=e)}_{\text{estimand}} \middle| R\right] = 0$$

$\Rightarrow$ $H(R)$ of dimension $|\mathcal{Y}|$ where $\mathbb{E}[H(R)W(o)]$ is full rank (invertible)

Some outcomes observed in the experiment

- Incorporate this information in the moments to improve efficiency

Some extensions

Non-binary outcomes:

- assuming no direct effects for simplicity

$$\mathbb{E}\left[\frac{1\{D=d,S=e\}}{P(D=d,S=e)} - \sum_{y\in\mathcal{Y}} \underbrace{\frac{1\{Y=y,S=o\}}{P(Y=y,S=o)}}_{\text{weight: } W_y(o)} \underbrace{P(Y(d)=y|S=e)}_{\text{estimand}} \Big| R\right] = 0$$

$\Rightarrow$ $H(R)$ of dimension $|\mathcal{Y}|$ where $\mathbb{E}[H(R)W(o)]$ is full rank (invertible)

Some outcomes observed in the experiment

- Incorporate this information in the moments to improve efficiency

Direct effect

- Identified if observe outcomes of some treated units (more)

Conclusions

- We study program evaluation with remote sensed variables
- We review pitfalls for practice and provide constructive identification
- No model specification for RSV is required and estimation/inference can use arbitrary ML algorithms
- We re-evaluate a large scale public policy program using satellite images and recover effects and SE with half of the sample

Conclusions

- We study program evaluation with remote sensed variables
- We review pitfalls for practice and provide constructive identification
- No model specification for RSV is required and estimation/inference can use arbitrary ML algorithms
- We re-evaluate a large scale public policy program using satellite images and recover effects and SE with half of the sample
- Many open questions for future research: spillover effects, high-dimensional outcomes, noisy measured treatment, ...

Thanks very much, questions?

Exact formula for bias of the surrogate

Specifically, we can write under the stated assumptions

$$\mathbb{E}[Y(d)|S=E] - \mathbb{E}[f(R)|S=E, D=d] =$$
$$\mathbb{E}[Y(d)|S=E] \cdot \int (1-w(d,r))P(R_i = r|Y_i = 1, S_i = e)dr$$

where

$$w(1,r) = \frac{P(Y(0)=1|S=E)P(R=r|D=1)}{P(Y(1)=1|S=E)P(R=r|D=0)}, \quad w(0,r) = \frac{1}{w(1,r)}$$

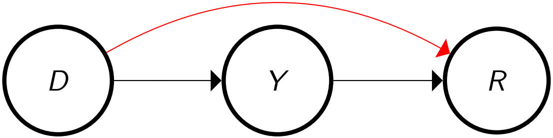
Taking differences, we obtain the bias for ATE of the form

$$\int \Big((1-w(1,r))\theta + \mu(0)(w(1,r)^{-1} - w(1,r))\Big)P(R_i = r|Y_i = 1, S_i = e)dr$$

where $\mu(d) = \mathbb{E}[Y(d)|S=E]$ (back)

Estimation with direct effects

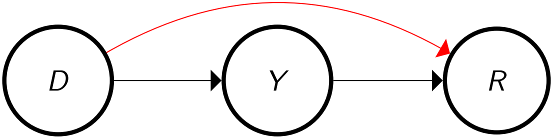
What if we also expect direct treatment effects?



$\Rightarrow$ Stability $S \perp R | Y, D$ implies stable direct effect (back)

Estimation with direct effects

What if we also expect direct treatment effects?



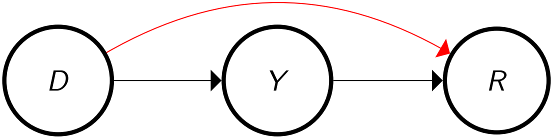
$\Rightarrow$ Stability $S \perp R | Y, D$ implies stable direct effect (back)

Thm (Informal) In the presence of direct effects

$$\mathbb{E}\left[\Delta_i(e) - \widetilde{\Delta}_i^0(o) - \theta\left(\widetilde{\Delta}_i^1(o) - \widetilde{\Delta}_i^0(o)\right)|R\right] = 0$$

Estimation with direct effects

What if we also expect direct treatment effects?



$\Rightarrow$ Stability $S \perp R | Y, D$ implies stable direct effect (back)

Thm (Informal) In the presence of direct effects

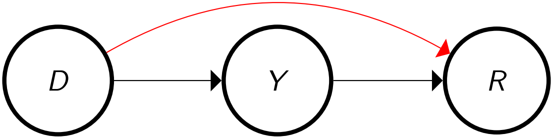
$$\mathbb{E}\left[\Delta_i(e) - \widetilde{\Delta}_i^0(o) - \theta\left(\widetilde{\Delta}_i^1(o) - \widetilde{\Delta}_i^0(o)\right) | R\right] = 0$$

where

$$\widetilde{\Delta}^y(o) = \frac{1\{Y = y, D = 1, S = o\}}{P(Y = y, D = 1, S = o)} - \frac{1\{Y = y, D = 0, S = o\}}{P(Y = y, D = 0, S = o)}$$

Estimation with direct effects

What if we also expect direct treatment effects?



$\Rightarrow$ Stability $S \perp R | Y, D$ implies stable direct effect (back)

Thm (Informal) In the presence of direct effects

$$\mathbb{E}\Big[\Delta_i(e) - \widetilde{\Delta}_i^0(o) - \theta\Big(\widetilde{\Delta}_i^1(o) - \widetilde{\Delta}_i^0(o)\Big) | R\Big] = 0$$

where

$$\widetilde{\Delta}^y(o) = \frac{1\{Y = y, D = 1, S = o\}}{P(Y = y, D = 1, S = o)} - \frac{1\{Y = y, D = 0, S = o\}}{P(Y = y, D = 0, S = o)}$$

$\Rightarrow$ Remove contribution of direct effect

Data description: Village (Shrid) level information

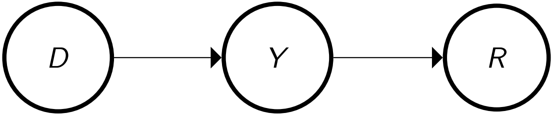
- Village boundaries through SHRUG project [Asher et al., 2021]
- Village per-capita consumption and poverty level with SECC 2012 (back)
- Village level nightlights with SHRUG
- Collect village level day-light satellites [MOSAIKS, Rolf et al., 2021]

Data description: Village (Shrid) level information

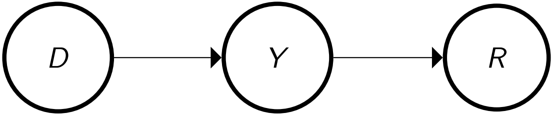
- Village boundaries through SHRUG project [Asher et al., 2021]
- Village per-capita consumption and poverty level with SECC 2012 (back)
- Village level nightlights with SHRUG
- Collect village level day-light satellites [MOSAIKS, Rolf et al., 2021]

	Outside Study	Control	Buffer Control	Treatment
Number of Shrids	2,260	853	2,931	2,276
Number mandals	105	44	136	111
Average pop	2,143	2,296	2,285	2,604
Urban area	0.002	0.001	0.003	0.004
Average male pop	0.512	0.508	0.506	0.508
Average female pop	0.489	0.492	0.495	0.493

Calibrated simulations. Draw

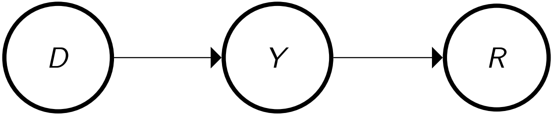


Calibrated simulations. Draw



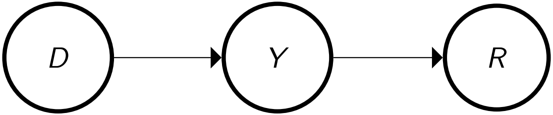
- Treatments D for n units independently
- $Y(1) \sim \mathrm{Bern}(\mathbb{E}_n[Y(1)] + \tau)$ ($Y(0) \sim \mathrm{Bern}(\mathbb{E}_n[Y(0)] + \tau)$)

Calibrated simulations. Draw



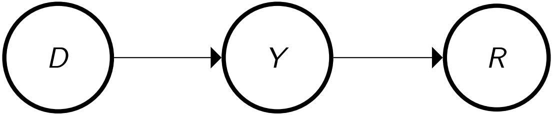
- Treatments D for n units independently
- $Y(1) \sim \mathrm{Bern}(\mathbb{E}_n[Y(1)] + \tau)$ ($Y(0) \sim \mathrm{Bern}(\mathbb{E}_n[Y(0)] + \tau)$)
- Draw $R(1)$ ($R(0)$) from pool of units with $Y = 1$ ($Y = 0$)

Calibrated simulations. Draw



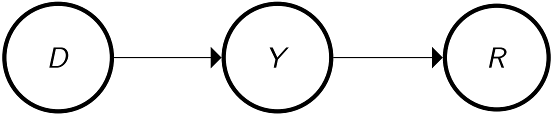
- Treatments D for n units independently
- $Y(1) \sim \mathrm{Bern}(\mathbb{E}_n[Y(1)] + \tau)$ ($Y(0) \sim \mathrm{Bern}(\mathbb{E}_n[Y(0)] + \tau)$)
- Draw $R(1)$ ($R(0)$) from pool of units with $Y = 1$ ($Y = 0$)
- Observe $Y = DY(1) + (1 - D)Y(0)$,

Calibrated simulations. Draw



- Treatments D for n units independently
- $Y(1) \sim \mathrm{Bern}(\mathbb{E}_n[Y(1)] + \tau)$ ($Y(0) \sim \mathrm{Bern}(\mathbb{E}_n[Y(0)] + \tau)$)
- Draw $R(1)$ ($R(0)$) from pool of units with $Y = 1$ ($Y = 0$)
- Observe $Y = DY(1) + (1 - D)Y(0), R = YR(1) + (1 - Y)R(0)$

Calibrated simulations. Draw



- Treatments D for n units independently
- $Y(1) \sim \mathrm{Bern}(\mathbb{E}_n[Y(1)] + \tau)$ ($Y(0) \sim \mathrm{Bern}(\mathbb{E}_n[Y(0)] + \tau)$)
- Draw $R(1)$ ($R(0)$) from pool of units with $Y = 1$ ($Y = 0$)
- Observe $Y = DY(1) + (1 - D)Y(0), R = YR(1) + (1 - Y)R(0)$
- Remove outcomes info for treated (back) individuals

Calibrated simulations: Surrogate vs RSV Method (RMSE)

(a) RMSE ($n = 2000$) (b) RMSE ($n = 3000$)