Remote Control: Debiasing Remote Sensing Predictions for Causal Inference

Matthew Gordon, Megan Ayers, Eliana Stone, Luke Sanford

Paris School of Economics, Reed College, Yale School of the Environment

ASSA 2026

Motivation: Satellite Data for Social Science

Remote sensing methods can help estimate effects of policy on land use changes

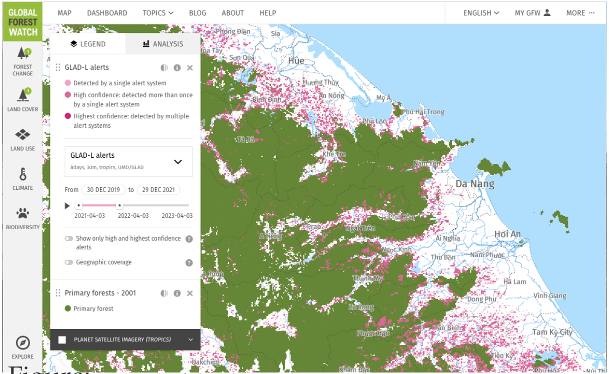


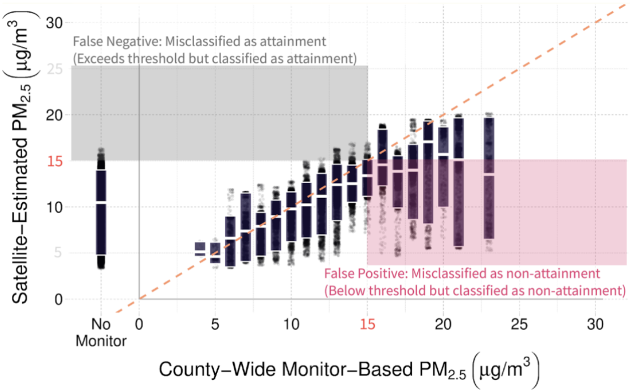
Figure: Hansen et al. (2013) `https://www.globalforestwatch.org/map/global/`

- Collecting data is expensive and time-consuming
- Machine learning methods + satellite data $\longrightarrow$ data sets of outcome variables with minimal labeling
- Global, high frequency, high resolution, free

Motivation: Satellite Data for Policy Analysis

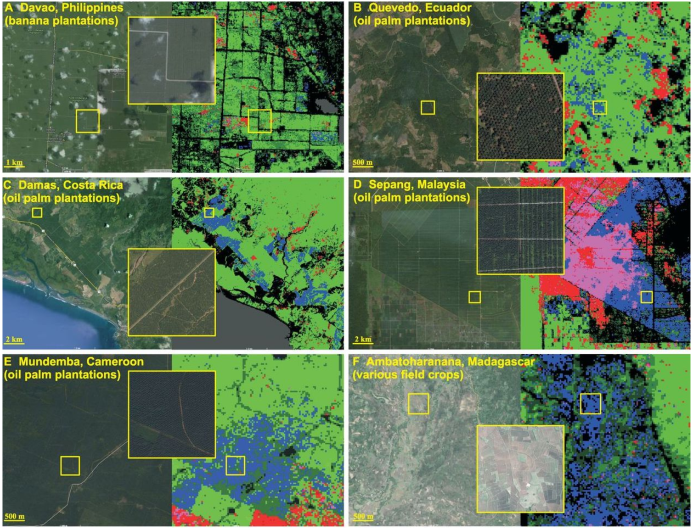
- Satellite data ↑ popularity in social sciences (Foster and Rosenzweig 2003, Henderson et al 2012, Burgess et al 2012)
- Well documented biases in Air Polluton[a]

[a] Fowlie, Rubin and Walker (2019)



Motivation: Satellite Data for Policy Analysis

- Satellite data ↑ popularity in social sciences (Foster and Rosenzweig 2003, Henderson et al 2012, Burgess et al 2012)
- Well documented biases in Air Polluton[a], Forest Cover[b]

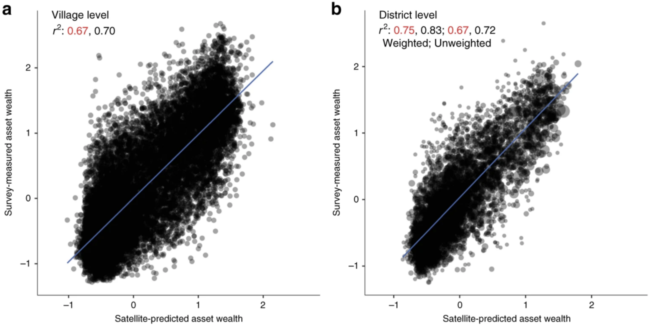


---

[a] Fowlie, Rubin and Walker (2019)
[b] Tropek et al. (2014)

Motivation: Satellite Data for Policy Analysis

- Satellite data ↑ popularity in social sciences (Foster and Rosenzweig 2003, Henderson et al 2012, Burgess et al 2012)
- Well documented biases in Air Polluton[a], Forest Cover[b], Wealth[c], and other variables can pose problems for causal inference



[a] Fowlie, Rubin and Walker (2019)
[b] Tropek et al. (2014)
[c] Ratledge et al. (2021)

Adversarial Debiasing

This Paper:

- We show how an ML algorithm called adversarial debiasing can generate 'unbiased' predictions for recovering treatment effects.
  - A general approach, with applications beyond remote sensing.
  - Potential efficiency gains over other bias correction methods.

Applications:

- Cross Sectional Simulations: Roads and forest cover in Africa.
- Panel: Gold mining and deforestation (Girard, Molina-Millán and Vic, 2025).
  - We create a new 'ground-truth' dataset of forest cover loss in Africa.
  - Correcting for measurement error shrinks loss estimates and widens confidence intervals.

Related Literature

Measurement Error in ML Predictions: (Ratledge et al., 2021; Alix-Garcia and Millimet, 2022; Angelopoulos et al., 2023; Torchiana et al., 2023; Proctor, Carleton and Sum, 2023; Kluger et al., 2025; Rambachan, Singh and Viviano, 2025; Carlson and Dell, 2025).

- Existing approaches 1) debias off-the-shelf predictions, and/or 2) typically rely on strong assumptions or knowledge of the sources of measurement error.

Algorithmic Justice, Machine Learning, Adversarial Debiasing: (Zhang, Lemoine and Mitchell, 2018; Chernozhukov et al., 2020; Liang, Lu and Mu, 2023; Arnold, Dobbie and Hull, 2024)

- Techniques to 'debias' ML algorithms with respect to race, gender, other protected characteristics.

Environment-Development Tradeoffs: (Foster and Rosenzweig, 2003; Benshaul-Tolonen, 2019; Alix-Garcia et al., 2013; Girard, Molina-Millán and Vic, 2025)

- Improved measurement of the effect of gold-mining on deforestation.

A Simple Setup:

We want to estimate:

$$Y_i = \alpha + \tau X_i + e_i$$

A Simple Setup:

We really estimate:

$$\widehat{Y_i} = \alpha + \tau X_i + e_i$$

A Simple Setup:

We really estimate:

$$\widehat{Y_i} = \alpha + \tau X_i + e_i$$

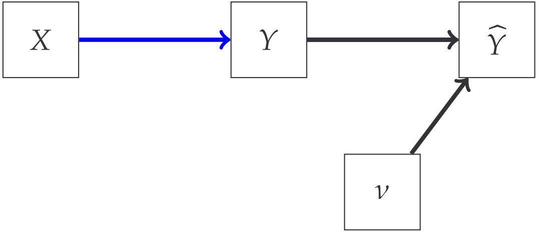
$$\widehat{Y_i} = Y_i + \nu_i$$

$$E[\widehat{\tau}] = \tau + \frac{cov(X, e)}{var(X)} + \frac{cov(X, \nu)}{var(X)}$$

Can even bias an RCT!

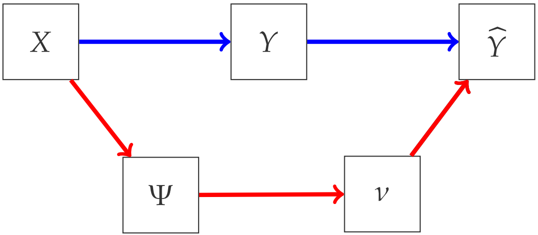
Prediction error and causal inference

- X: Wealth Shock, Y: Forest cover



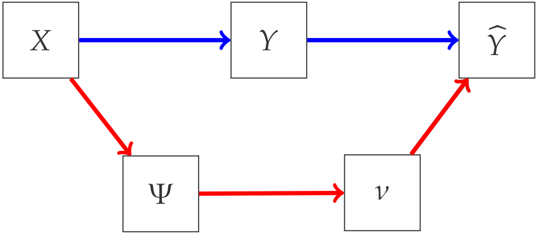
Prediction error and causal inference

- X: Wealth Shock, Y: Forest cover
- Problem: ψ is irrigated cropland, more often misclassified as tree cover.



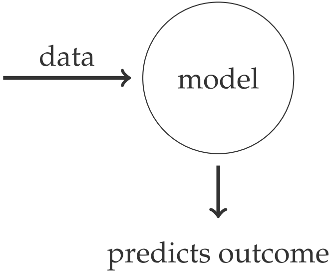
Prediction error and causal inference

- X: Wealth Shock, Y: Forest cover
- Problem: ψ is irrigated cropland, more often misclassified as tree cover.

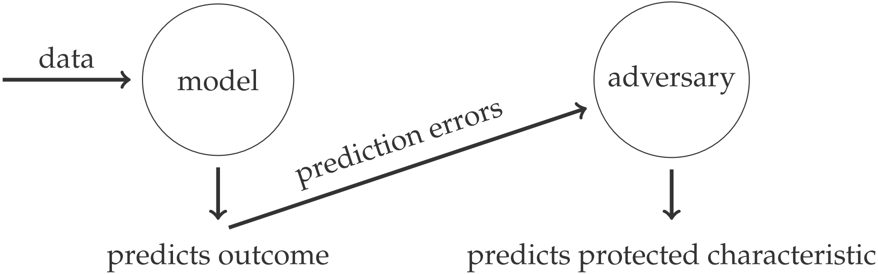


Question: Can we generate predictions such that $cov(X, \nu) = 0$?

Introduction to adversarial debiasing

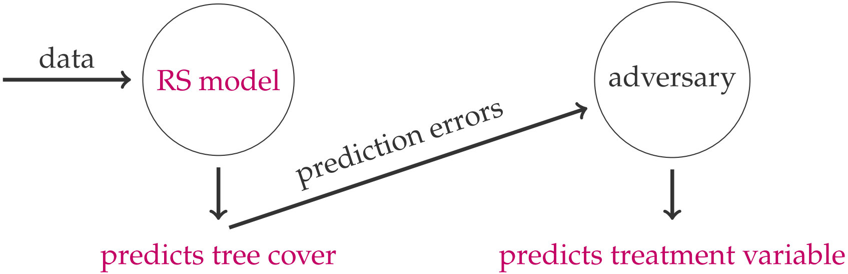


Introduction to adversarial debiasing



Used in computer science for making sure e.g. loan applications/bail decisions/college admissions do not discriminate on the basis of e.g. race or gender (Zhang et al. 2018)

Introduction to adversarial debiasing



Used in computer science for making sure e.g. loan applications/bail decisions/college admissions do not discriminate on the basis of e.g. race or gender (Zhang et al. 2018)

Adversarial debiasing formally

We want to choose model weights ω according to:

$$\begin{gathered}\omega^* = \arg\min_{\omega} L_p(\widehat{Y}(\omega), Y, k) \\ \text{such that } Cov(X, Y - \widehat{Y}(\omega)) = 0.\end{gathered} \tag{1}$$

Adversarial Debiasing (Zhang, Lemoine and Mitchell, 2018):

$$\begin{gathered}\min_{\omega} \max_{\gamma} \left\{ L_p \left(\hat{Y}(\omega), Y, k \right) - \alpha \cdot L_a \left(X, Y, \widehat{Y}(\omega), \gamma, k \right) \right\} \\ \text{such that } \gamma \in \text{argmin } L_a(X, Y, \widehat{Y}(\omega), \gamma, k)\end{gathered} \tag{2}$$

Unbiased Predictions

Consider an adversary that is a linear regression of treatment on prediction errors (e.g. L_a is MSE):

$$\nu_i = \gamma X_i + \epsilon_i \tag{3}$$

Unbiased Predictions

Consider an adversary that is a linear regression of treatment on prediction errors (e.g. L_a is MSE):

$$\nu_i = \gamma X_i + \epsilon_i \tag{3}$$

Constraint ensures that γ equals expected bias in τ:

$$\widehat{\gamma} = \frac{cov(X, \nu)}{var(X)} \tag{4}$$

Isomorphic to objective 1 when α is chosen to be Lagrangian multiplier on covariance constraint.

Efficiency

$$\widehat{\gamma} = \frac{cov(X, v)}{var(X)} \tag{5}$$

Intuitively, for a given accuracy, adversary MSE is maximized when $\widehat{\gamma} = 0$. Model will attempt to choose ω to achieve this.

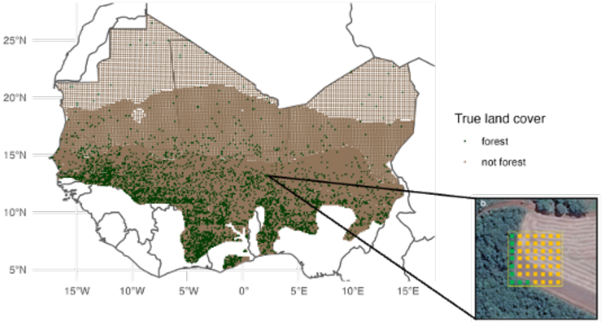
- In contrast, predict-then-debias (PTD) approaches estimate γ holding predictions fixed.

Relative efficiency depends on tradeoff between bias and accuracy:

- Efficiency of PTD is constrained by variance of $\widehat{\gamma}$. Depends on size of labeled set and variance of original prediction errors.
- If adversarial model can learn predictions that are unbiased without sacrificing too much accuracy, sampling error from estimating γ in small labeled set dominates.

Simulations: Can we recover τ using ML predictions of Y?

- Bastin (2017) 23,000 hand-labeled points on dryland forest in W. Africa



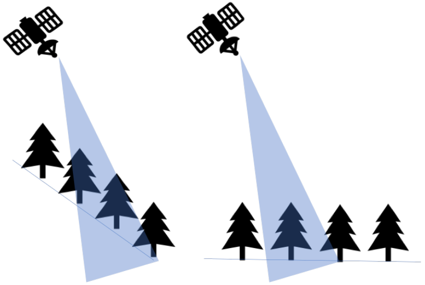
- Train baseline model, adversarial model on Landsat 7 data, estimate τ and compare to ground truth.

Simulation: $\tau = 0$

- $\psi \sim$ Poisson (think e.g. slope)
- $p(X)$ decreasing in ψ (think e.g. infrastructure)
- Y: % Forest cover - randomly draw points and associated satellite data

Simulation: $\tau = 0$

- $\psi \sim$ Poisson (think e.g. slope)
- $p(X)$ decreasing in ψ (think e.g. infrastructure)
- Y: % Forest cover - randomly draw points and associated satellite data
  - One catch: if $\psi > 0$, we make the points look 'greener'



Simulation: $\tau = 0$

- $\psi \sim$ Poisson (think e.g. slope)
- $p(X)$ decreasing in ψ (think e.g. infrastructure)
- Y: % Forest cover - randomly draw points and associated satellite data
  - One catch: if $\psi > 0$, we make the points look 'greener'

- Since high ψ points are less common, standard ML model learns that green usually means trees
- Debiased model notices these errors are correlated with X, does worse on low ψ points, better on high ψ points

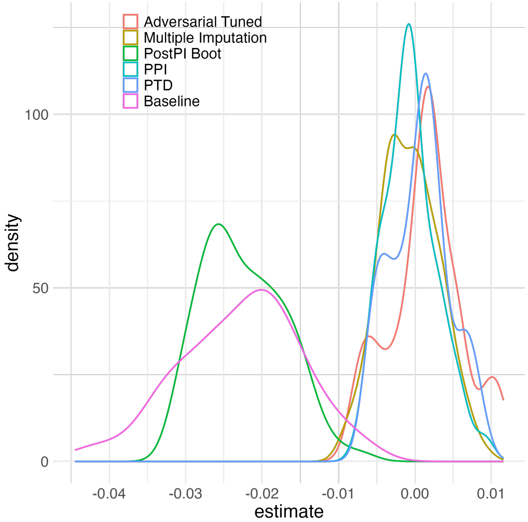
Simulation: $\tau = 0$

- $\psi \sim$ Poisson (think e.g. slope)
- $p(X)$ decreasing in ψ (think e.g. infrastructure)
- Y: % Forest cover - randomly draw points and associated satellite data
  - One catch: if $\psi > 0$, we make the points look 'greener'

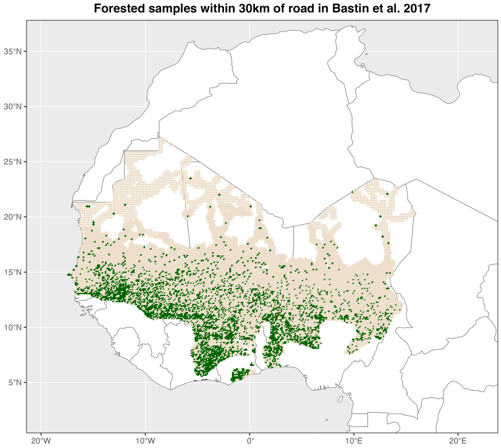
- Since high ψ points are less common, standard ML model learns that green usually means trees
- Debiased model notices these errors are correlated with X, does worse on low ψ points, better on high ψ points
  - This is despite not knowing ψ!

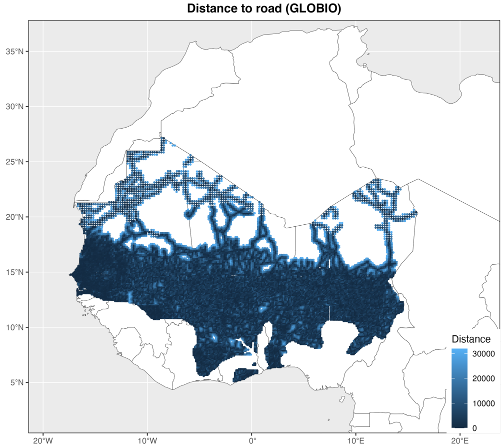
Simulation: Comparing methods

- Bootstrapped distributions from 100 training runs, 10k points
- Baseline model biased, most other methods perform well



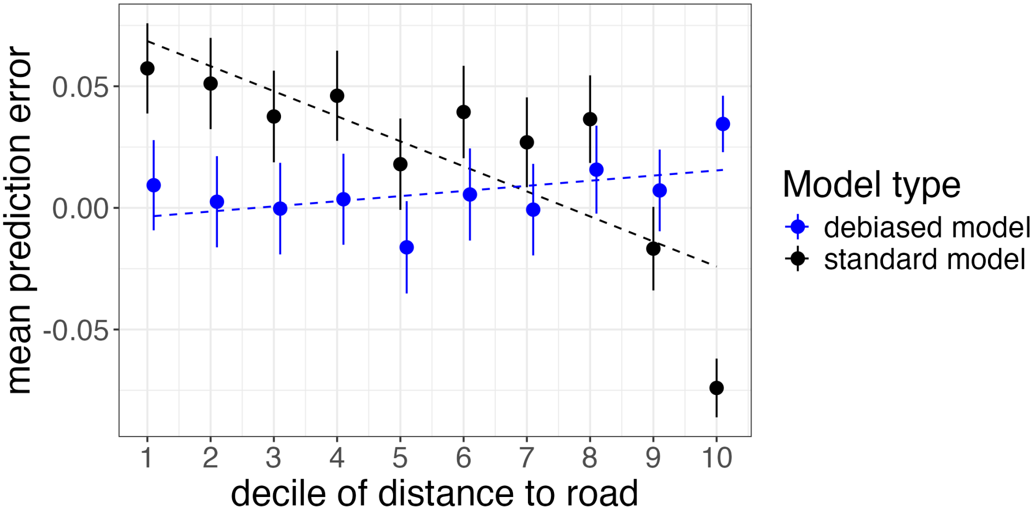
Descriptive Application: Roads and Forest Cover





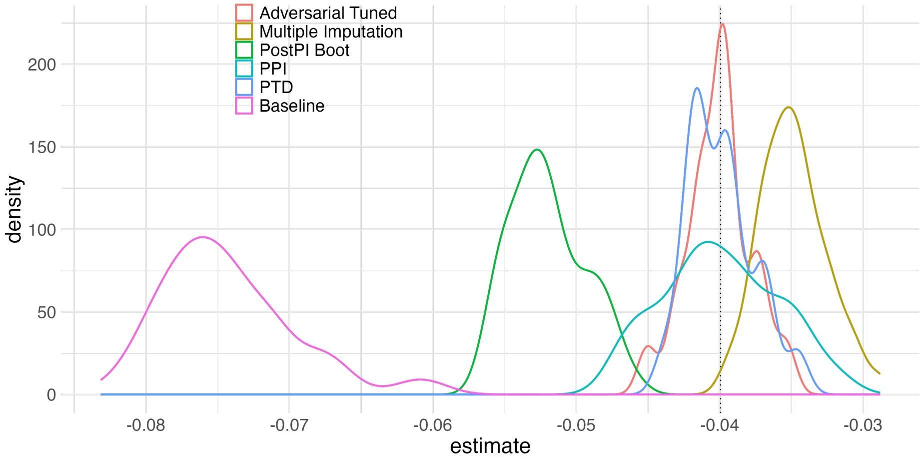
Estimate simple cross-sectional relationship using standard and debiased model predictions.

Covariance between X and ν



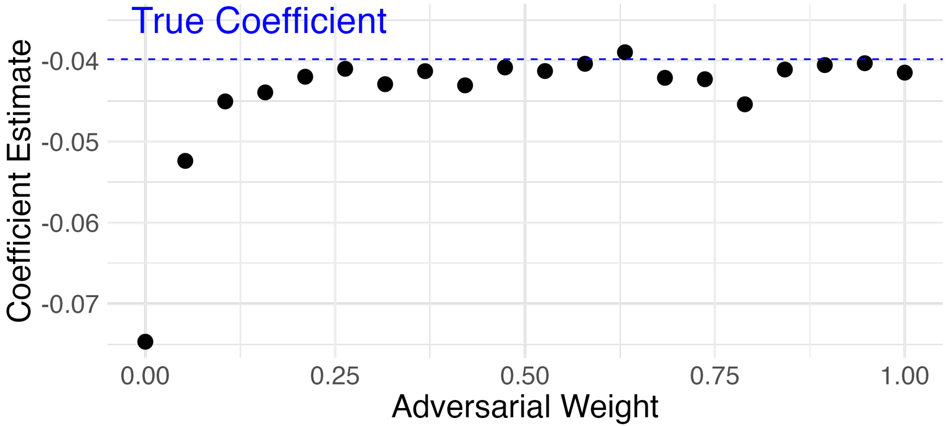
Descriptive Application: Assessing Bias

Figure: Bootstrapped Coefficient Distributions



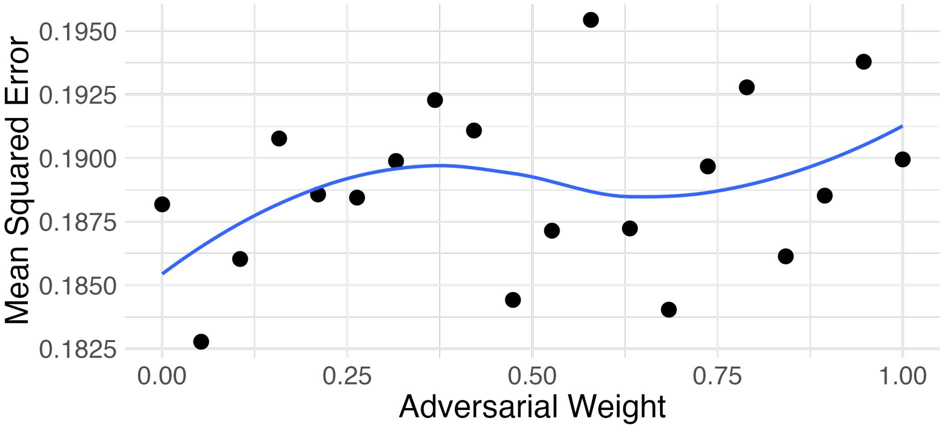
Tuning α: Precision-bias tradeoff?

Figure: Primary Model: Logistic Regression



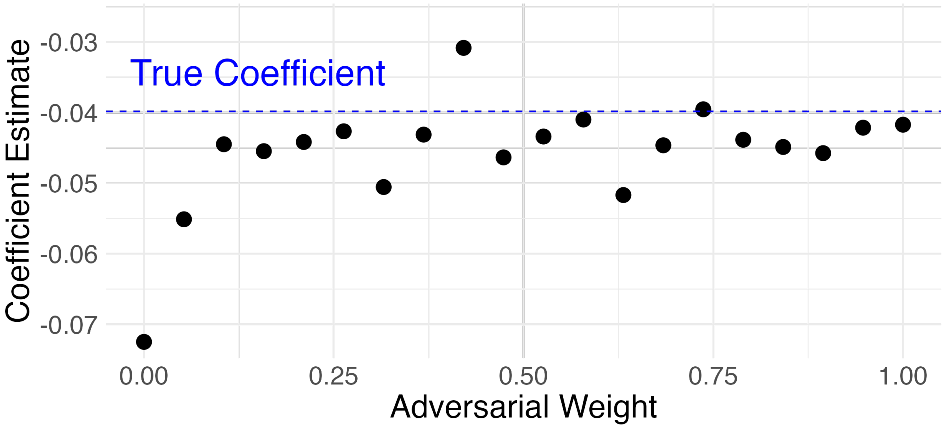
Tuning α: Precision-bias tradeoff?

Figure: Primary Model: Logistic Regression



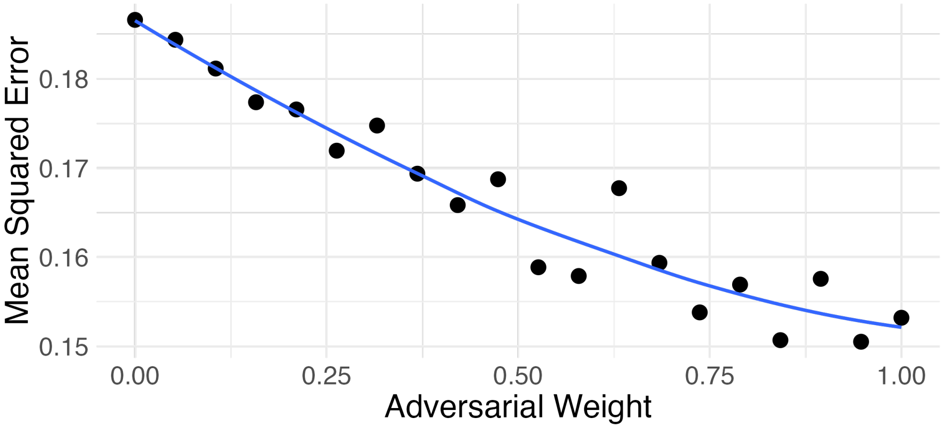
Tuning α: Precision-bias tradeoff?

Figure: Primary Model: Deep Neural Net



Tuning α: Precision-bias tradeoff?

Figure: Primary Model: Deep Neural Net



Causal Application: Artisanal Gold Mining and Deforestation

Primary specification in Girard, Molina-Millán and Vic (2025):

$$\widehat{\widehat{Y}}_{cjt} = \beta_1 G_c \times P_{t-1} + \mu_c + \lambda_{jt} + \epsilon_{ct}, \tag{6}$$

- Finds changes in int'l gold prices result in more deforestation in gold suitable areas using Hansen et al. (2013).
- Benshaul-Tolonen (2019): gold rush increased wealth, health.
- Foster and Rosenzweig (2003) argues that increases in wealth in India $\rightarrow$ increased forest cover due to increased demand for forest products.

Causal Application: Artisanal Gold Mining and Deforestation

Primary specification in Girard, Molina-Millán and Vic (2025):

$$\widehat{\bar{Y}}_{cjt} = \beta_1 G_c \times P_{t-1} + \mu_c + \lambda_{jt} + \epsilon_{ct}, \tag{6}$$

- Finds changes in int'l gold prices result in more deforestation in gold suitable areas using Hansen et al. (2013).
- Benshaul-Tolonen (2019): gold rush increased wealth, health.
- Foster and Rosenzweig (2003) argues that increases in wealth in India → increased forest cover due to increased demand for forest products.

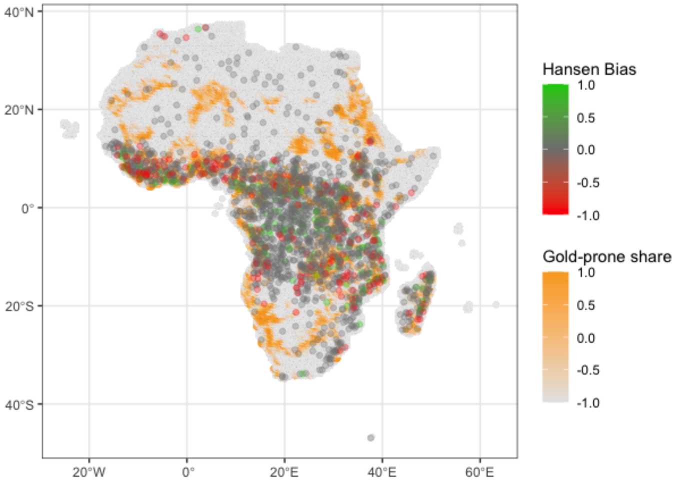
Problem: No good time-series ground-truth data on deforestation.

Ground Truth Data Collection: Active Sampling

Google Earth Pro Historical Imagery

Measurement Error in GFW Data

- Train small model on limited data.
- Use model to predict out of sample
- Select 2000 pts proportionate to model uncertainty
- See Zrnic and Candès (2024) and Gordon and Papp (Forthcoming)



Debiased Results

	$\widehat{\bar{Y}}_g$ (1)	$\widehat{Y}_g$ (2)	$\widehat{Y}_i$ (3)	ν (4)	$\widehat{Y}_i^D$ (5)
gold suitable × price	0.3033***	9.362×10^{-4}***	0.0018***	0.0013	1.01×10^{-4}***
	(0.0448)	(1.384×10^{-4})	(4.109×10^{-4})	(0.0050)	(1.77×10^{-5})
FE: grid cell	X	X	X	X	X
FE: year^country	X	X	X	X	X
Num.Obs.	192 348	192 348	900 000	36 000	899 442
Kluger et al. (2025) 95% CI			[-0.0070, 0.0083]		[-0.0049, 0.0080]

Regression results of equation 6. Columns 1-3 use Hansen et al. (2013) data, column 4 uses measurement error as the dependent variable, and column 5 uses our debiased predictions. All regressions cluster standard errors by grid cell following Girard, Molina-Millán and Vic (2025).

Takeaways

- Caution when using proxies for outcome variables
  - Issue is not unique to remote sensing! Many uses of machine-learned outcomes as variables in other models (eg. text-based outcomes)
- Variety of methods will debias estimates:
  - Adversarial debiasing may improve efficiency, but important to still use a post-prediction debiaser.
- A simple unbiased model can be better than a more powerful more accurate model
- Need for more ground-truth data collection!

Thank you!

matthew.gordon@psemail.eu