Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs

Luna Yue Huang*

Joint with Solomon Hsiang and Marco Gonzalez-Navarro

UC Berkeley

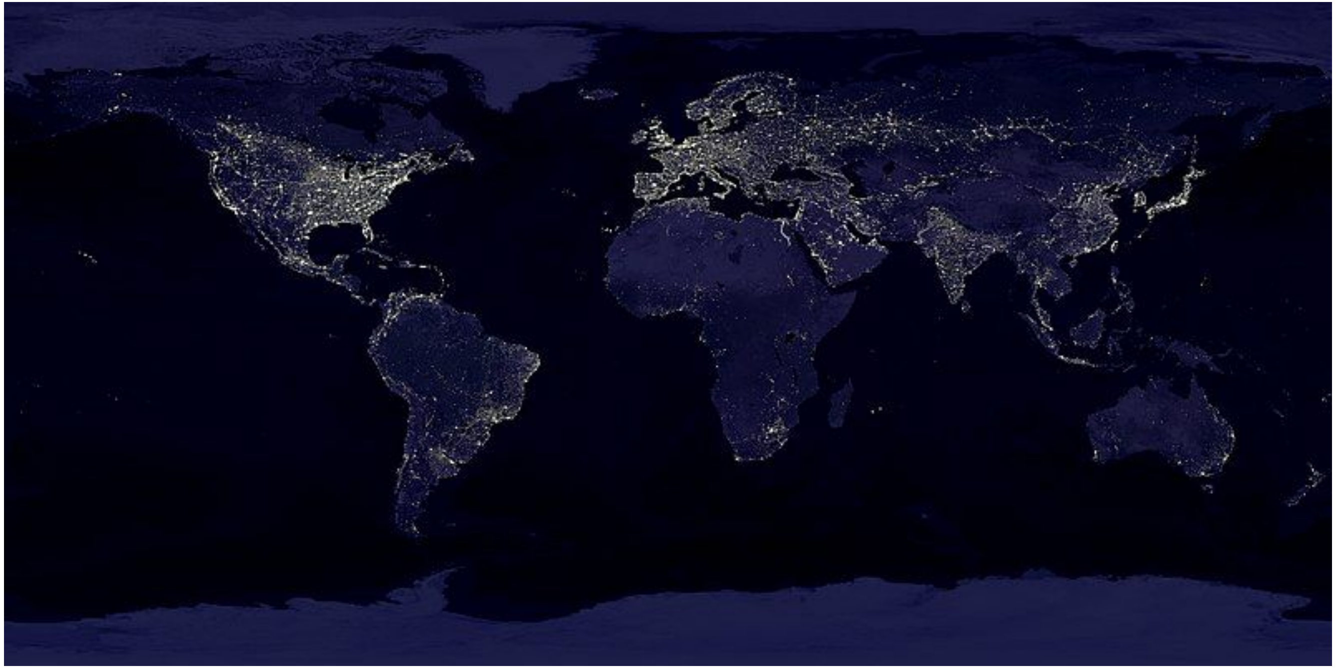
FAO Presentation | June 28, 2021

Can we use remote sensing data to bring down the costs of doing program evaluation in development economics?

- Household surveys are
  - very expensive - $18–300 per household
  - very often disrupted by
    - political unrest
    - epidemic / pandemic
    - other unanticipated events
- What if we can do impact evaluation without going to the field?
  - Observe the quantity and quality of physical assets, in particular houses that people live in
  - with high-resolution daytime satellite imagery and deep learning
  - $0.006 per household

Motivation

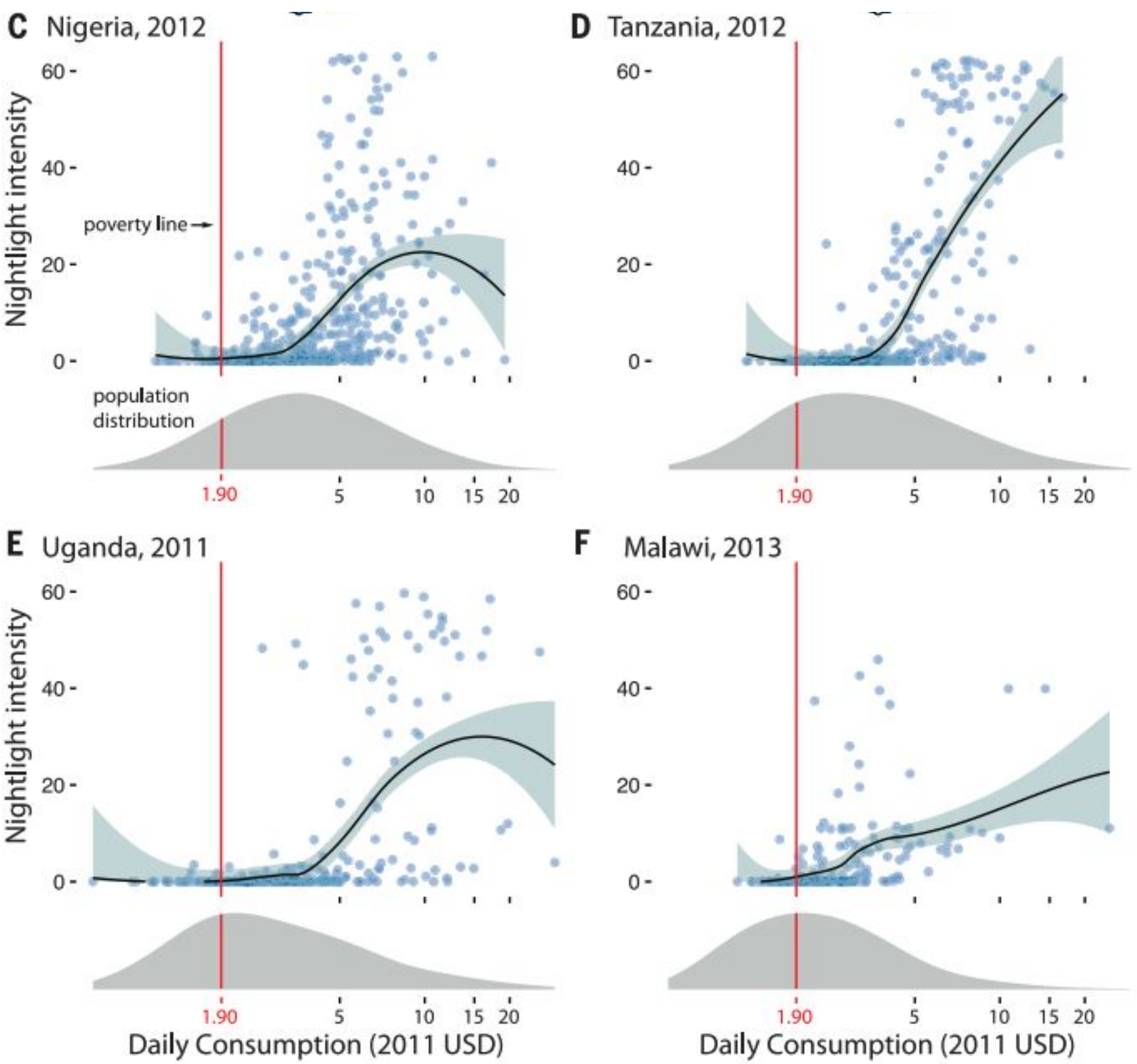
- Prominent literature on night light (Henderson et al. 2012)



Motivation

- Poor sensitivity of night light in low-income and rural contexts (Jean et al. 2016)

But that's arguably what development economists are most interested in!



Motivation

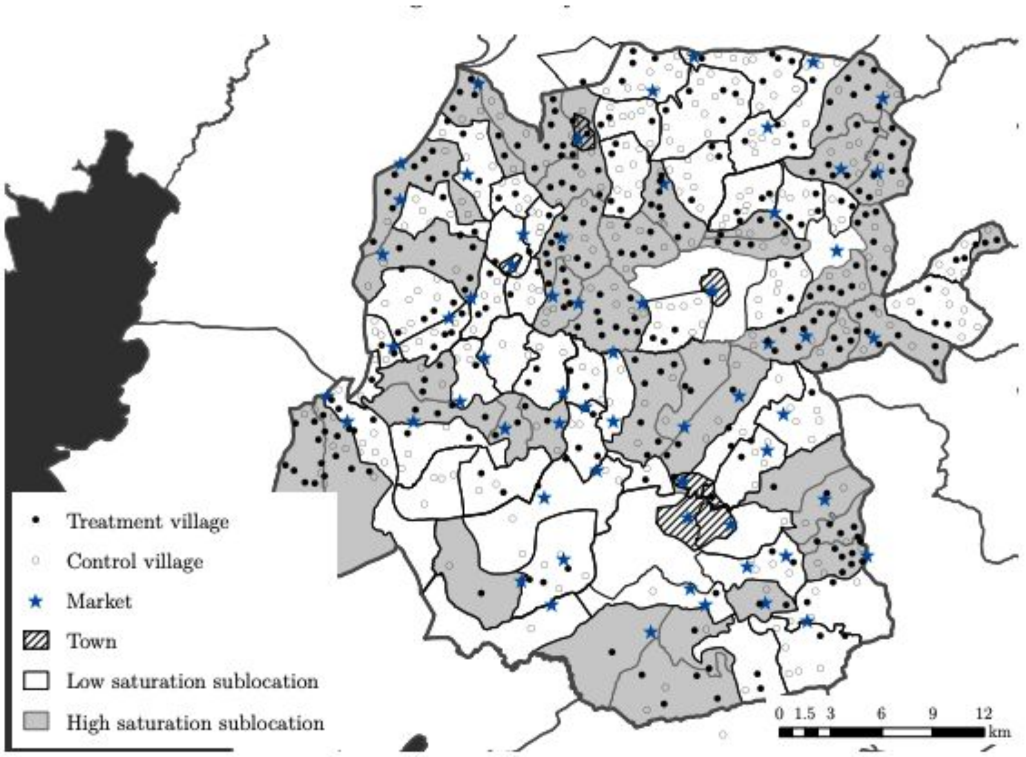
- What if we observe housing? It may be a better outcome ...
  - Sensitive even in communities with low electrification rates
  - Can be precisely measured in daytime satellite imagery (high resolution)
  - Accounts for a much larger share of expenditure than electricity

This paper

- proposes a framework for conducting impact evaluation with satellite imagery and deep learning
- evaluates the GiveDirectly randomized controlled trial in rural Kenya as a proof of concept
- shows that
    - a state-of-the-art deep learning model can generate accurate **housing quality** measurements at scale
    - statistically significant and economically sizable treatment effects can be observed from remotely sensed outcome variables
    - when certain assumptions hold, we can recover the overall effects of the intervention on household wealth with a simple scaling exercise

What is GiveDirectly?

- A large scale randomized controlled trial in rural Kenya in 2014-2017 (Egger et al. 2019)
- Distributes unconditional cash transfers to rural households (~$1,000 per household)
- About ⅓ of the households in the treatment villages are eligible to get the transfer



Why GiveDirectly?

To be evaluated with satellite imagery:

- Plausible impacts on physical assets (especially housing), which is observable from satellite imagery
- Spatially explicit random variation in treatment intensity
- Recent intervention (2014-2017)

As a validation exercise:

- Randomized experiment - clean set up
- We know what the true treatment effects are (Egger et al. 2019)

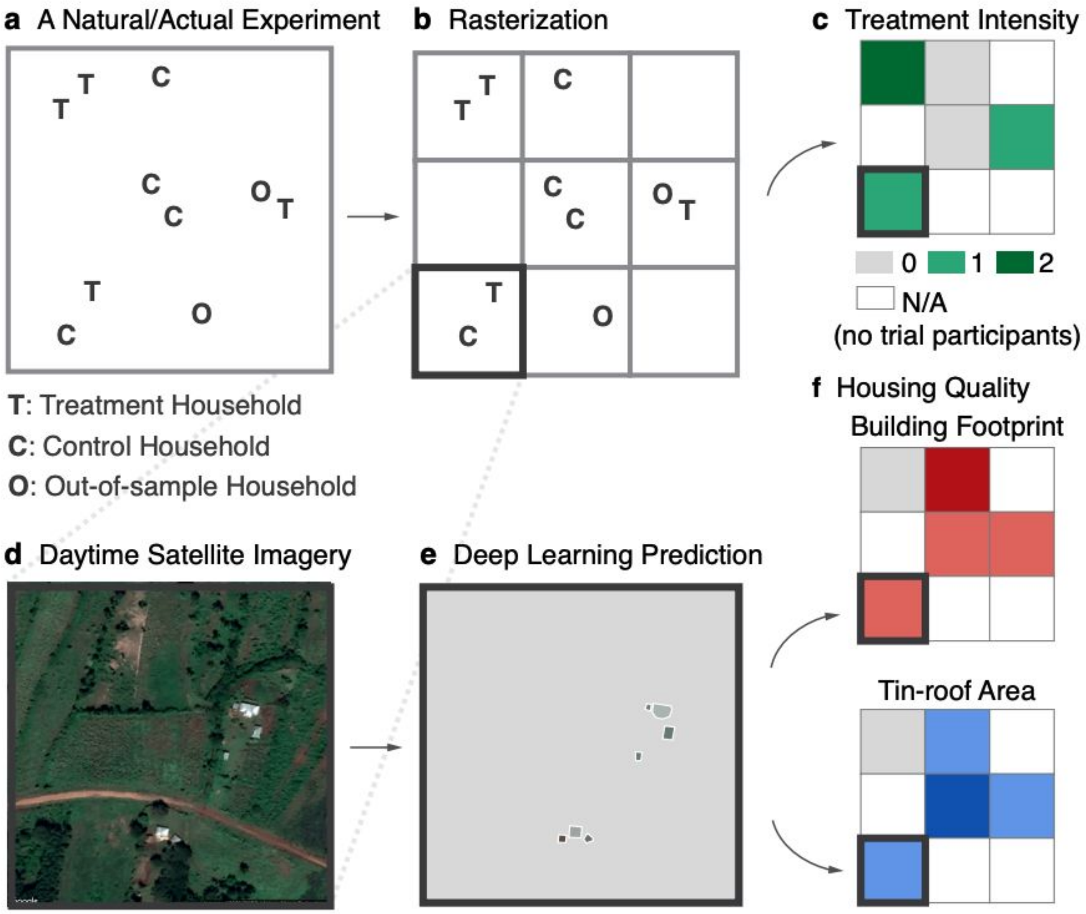
Table 1: Expenditures, Savings and Income

	(1)	(2)	(3)
	Treated Households		Untreated Households
	1(Treat village) Reduced form	Total Effect IV	Total Effect IV
Panel A: Expenditure			
Household expenditure, annualized	292.98***	343.34***	333.75***
	(60.09)	(112.02)	(123.23)
Non-durable expenditure, annualized	174.99***	211.90**	288.54***
	(55.41)	(96.76)	(111.45)
Food expenditure, annualized	71.61*	138.58**	132.85**
	(36.93)	(66.76)	(58.58)
Temptation goods expenditure, annualized	6.51	4.47	−0.71
	(5.79)	(9.18)	(6.50)
Durable expenditure, annualized	95.18***	106.29***	8.41
	(12.64)	(21.44)	(12.50)
Panel B: Assets			
Assets (non-land, non-house), net borrowing	179.99***	164.00***	111.39
	(27.46)	(54.02)	(90.18)
Housing value	378.36***	377.27***	67.75
	(26.65)	(40.47)	(311.56)
Land value	−75.77	179.03	478.50
	(206.47)	(332.75)	(512.54)

Methods + Results

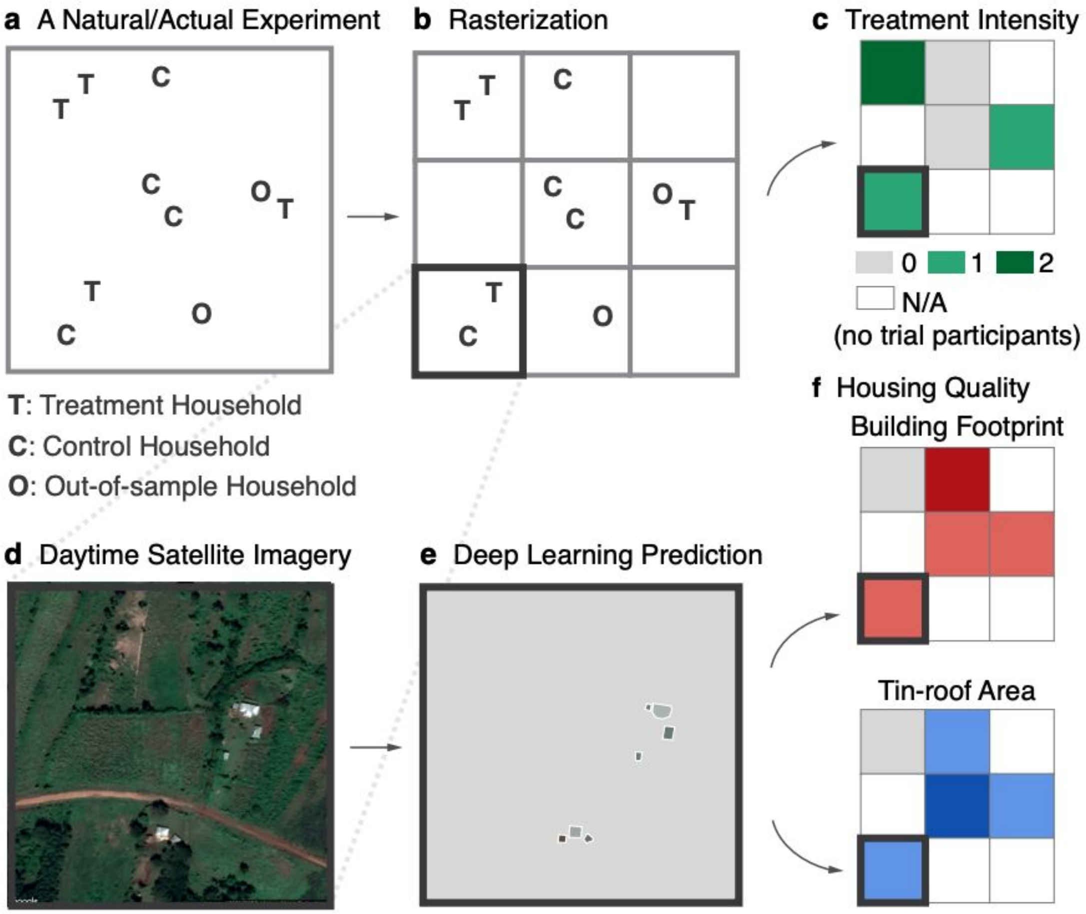
How to measure treatment intensity?

- Data source: GiveDirectly baseline census (~65,000 household geo-location + eligibility status + treatment status)
- Define treatment intensity as the no. of households who ultimately received transfers from GiveDirectly
- Equivalently, the amount of cash infusion (in $1,000) into a given grid cell



How to measure housing quality?

- Input images: Google Static Map, 2019 (submeter imagery)
- Deep learning model: Mask R-CNN for instance segmentation: binary pixel mask -> polygon -> “representative” color -> classify roof type



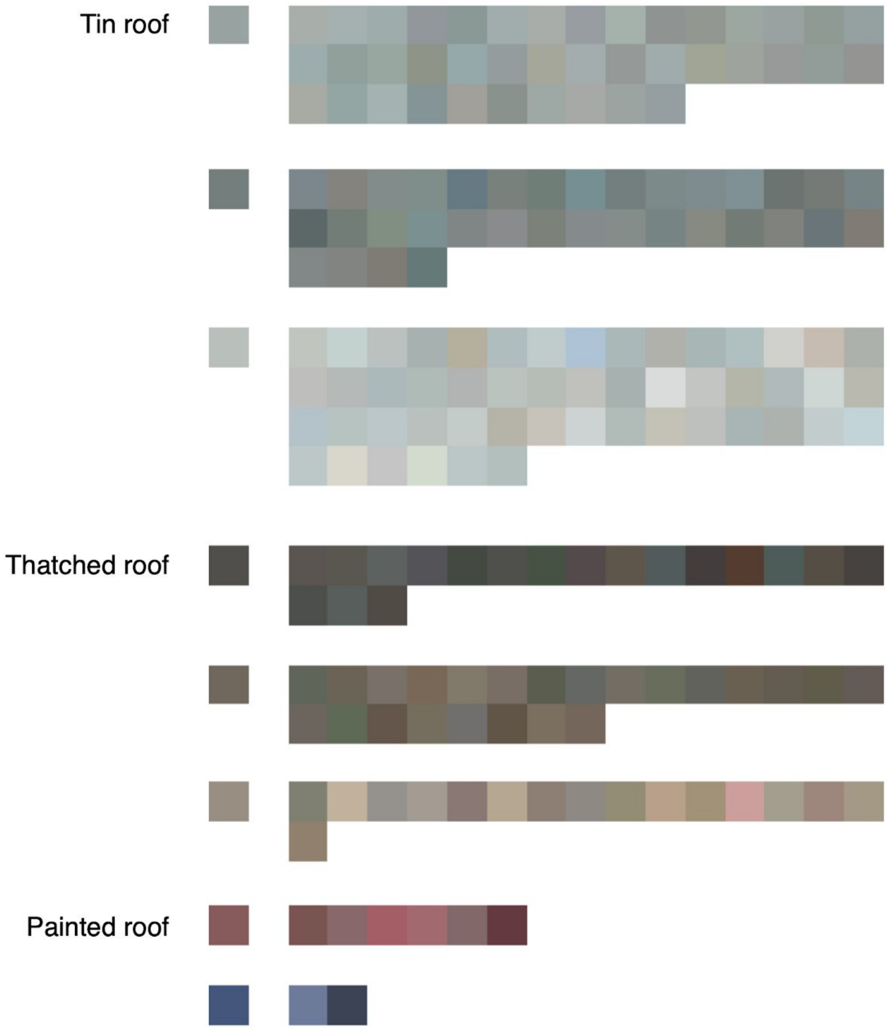
Deep learning model performance (precision/recall: ~80%)

10 random samples

Classify roof types

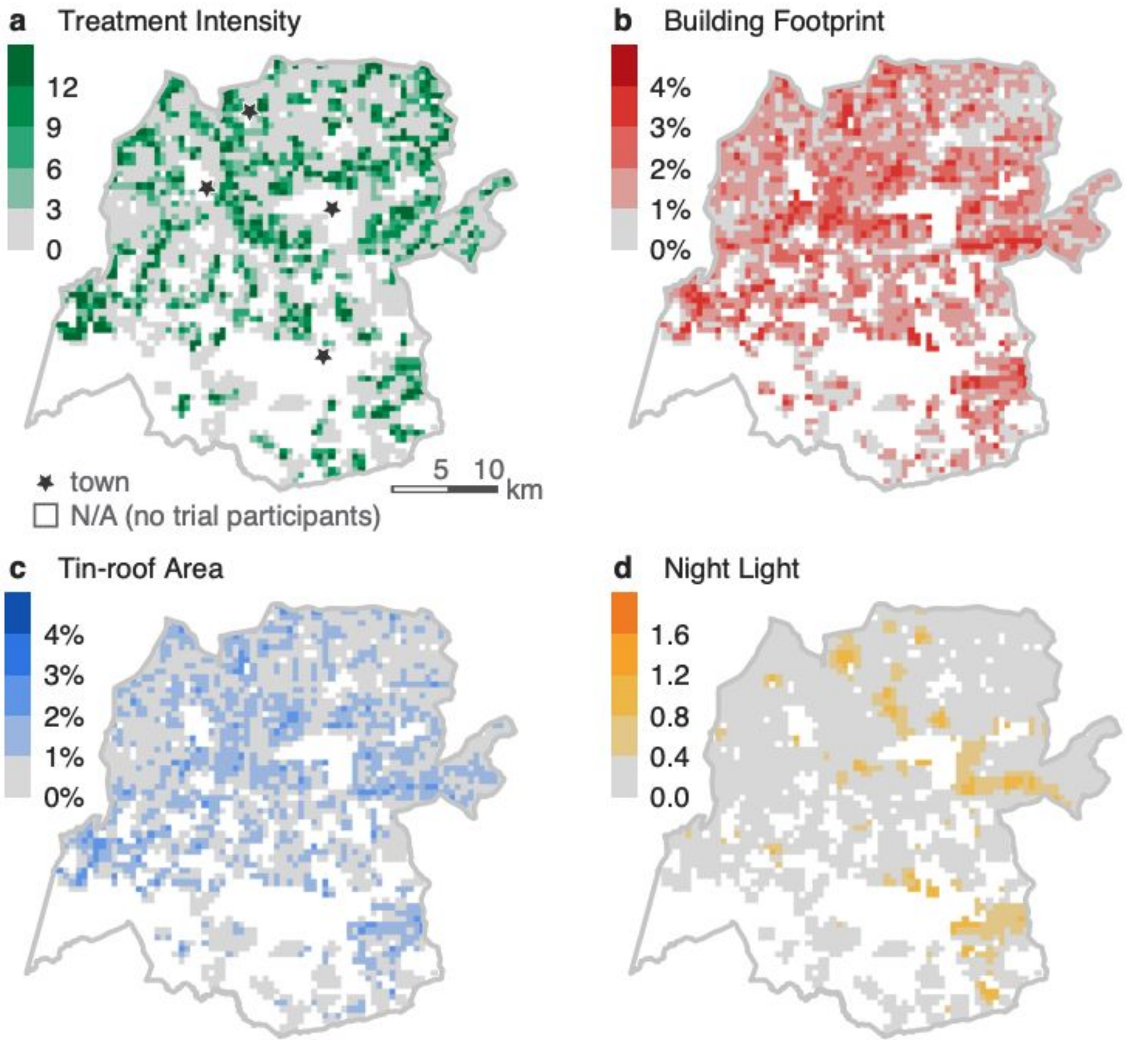
- Tin roofs are of higher quality and more durable than thatched roofs

- We use tin roofs as a proxy for high quality housing asset



Treatment intensity and outcomes in the study area

- Night light: less variation in rural areas



Main Result

remotely sensed outcomes

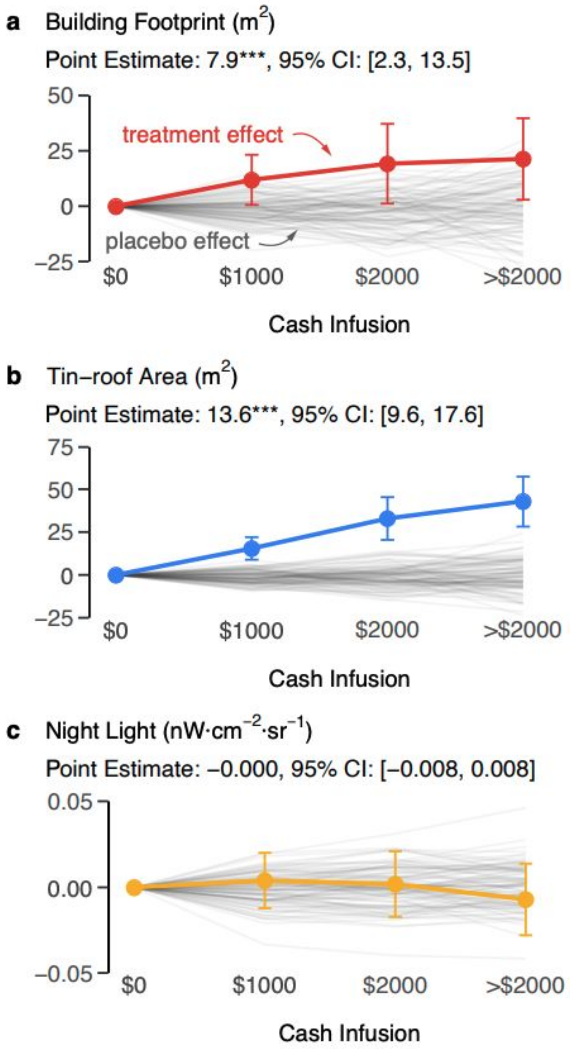
treatment intensity

no. of eligible households

$$y_i = \sum_{k \in K} \tau_k \mathbf{1}\{x_i = k\} + \sum_{m \in M} \beta_m \mathbf{1}\{e_i = m\} + \epsilon_i$$

*each observation = one pixel (0.001x0.001 degree, or roughly 100m by 100m)

Placebo effects are estimated by re-simulating the treatment/control group randomization with the original trial design.



= 85 sq ft

= 146 sq ft

Estimating the overall effects of the trial

- Engel curve:

$$Q_{hp} = \alpha_p + \beta_p W_h + \epsilon_{hp}$$

- One can prove that

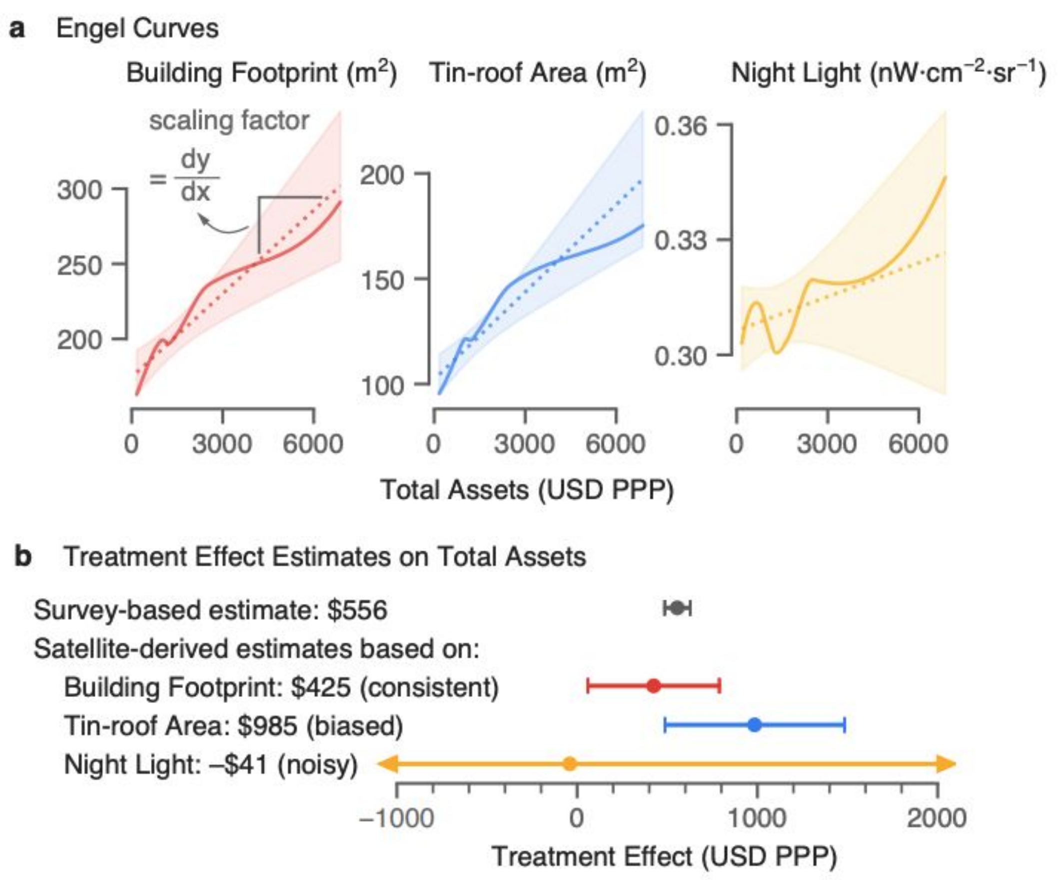
$$\hat{\tau}_W = \hat{\tau}_{Q_p} / \hat{\beta}_p$$

scaling factor

Effects on wealth

Effects on housing

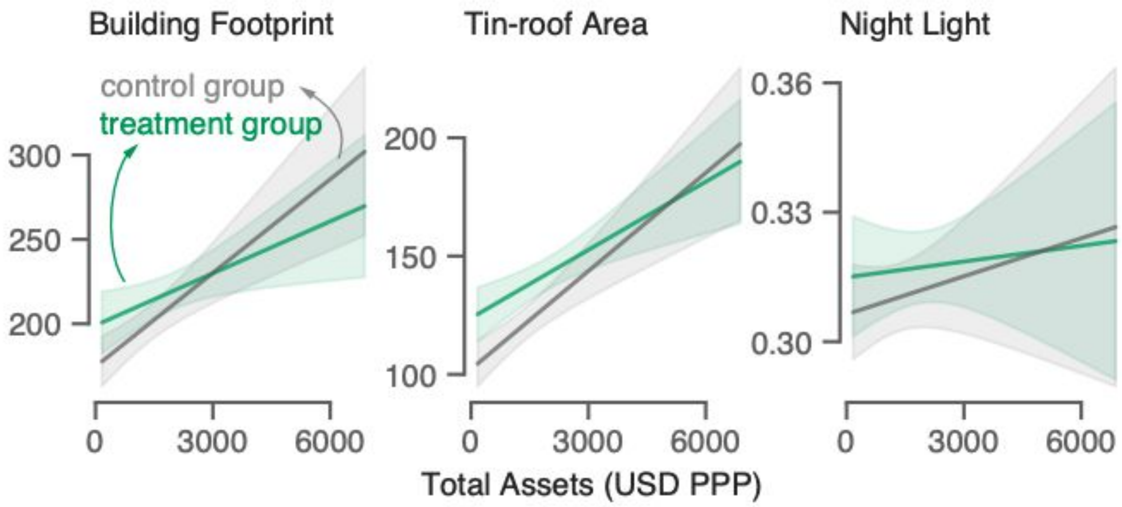
- **...if the engel curve does not change in response to the treatment**



The Engel curve for high quality (tin) roofs changed in response to the treatment

Behavioral change:

Households are eligible for the GiveDirectly cash transfer if they live in thatched roof houses. They might have psychologically tied the transfer to roof upgrading (as if this is a “labelled” cash transfer).



These results highlighted the danger of using black box machine learning predictions for impact evaluation

- Poverty mapping with satellite imagery and machine learning, developed mostly for precisely targeting international aid (Jean et al. 2016, Yeh et al. 2020, Blumenstock et al. 2015, Blumenstock 2016, Aiken et al. 2020, and many others)
  - Does not specify a source of information (electrification, housing, infrastructure, etc.), let the model use all the information available
- If used for impact evaluation, some information will be “tainted” - i.e. directly changed by the intervention; imagine:
  - Using night lights to evaluate an electrification campaign
- **More interpretable machine learning predictions directly measuring housing consumption or other assets can be used for impact evaluation**

Thank you!

yue_huang@berkeley.edu

We thank Edward Miguel, Jeremy Magruder, Ben Faber, Marshall Burke, Joshua Blumenstock, Supreet Kaur, Ethan Ligon, Elisabeth Sadoulet, Alain De Janvry, Aprajit Mahajan, the participants in the AGU Fall Meeting 2019 (Session GC34C), the UC Berkeley Trade Lunch, Development Workshop, Development Lunch, and Good Data Seminar for feedback. We thank Edward Miguel, Michael Walker, Dennis Egger, Johannes Haushofer, Paul Niehaus, and the rest of the GiveDirectly team, for generously sharing the dataset with us and responding to our inquiries.