Poverty Mapping in the Age of Machine Learning

Paul Corral, Heath Henderson, and Sandra Segovia

August 27, 2024

Introduction

- Recent developments in poverty mapping have departed from standard practice in important ways.

Introduction

- Recent developments in poverty mapping have departed from standard practice in important ways.
- In brief, standard practice consists of two basic steps:
  1. Fit a parametric statistical model by regressing a survey-based welfare measure on some area-level covariates, typically census aggregates.
  2. Predict the welfare measure for all geographic entities using the census aggregates.

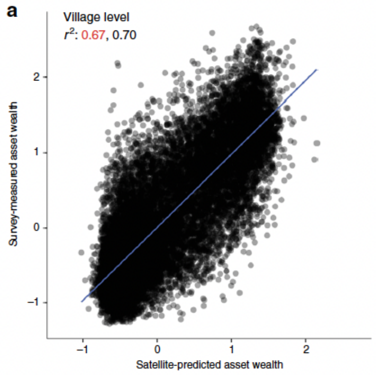
Introduction

- Recent developments in poverty mapping have departed from standard practice in important ways.
- In brief, standard practice consists of two basic steps:
  1. Fit a parametric statistical model by regressing a survey-based welfare measure on some area-level covariates, typically census aggregates.
  2. Predict the welfare measure for all geographic entities using the census aggregates.
- In contrast to this "traditional" approach, the "modern" approach:
  1. Relies on non-parametric, machine-learning methods rather than parametric models
  2. Tends to rely more on non-traditional, remotely-sensed covariates in the estimation and prediction stage

Introduction

- Recent developments in poverty mapping have departed from standard practice in important ways.
- In brief, standard practice consists of two basic steps:
  1. Fit a parametric statistical model by regressing a survey-based welfare measure on some area-level covariates, typically census aggregates.
  2. Predict the welfare measure for all geographic entities using the census aggregates.
- In contrast to this "traditional" approach, the "modern" approach:
  1. Relies on non-parametric, machine-learning methods rather than parametric models
  2. Tends to rely more on non-traditional, remotely-sensed covariates in the estimation and prediction stage
- In what follows, we'll discuss this modern approach and evaluate its performance against a credible ground truth.

An Example



"These results exceed performance in earlier work on a similar task using high-resolution imagery or mobile phone data as input, and match or exceed benchmarks for in-country performance from geostatistical models used to predict health outcomes, standard of living, and housing quality in Africa" (Yeh et al. 2020)

An Example

- The standard method for assessing performance calculates R^2 as follows:

$$y_i^d = \alpha + \delta y_i^p + \zeta_i \quad \Rightarrow \quad R_d^2 = \frac{\sigma_d^2 - \sigma_\zeta^2}{\sigma_d^2}$$

An Example

- The standard method for assessing performance calculates R^2 as follows:

$$y_i^d = \alpha + \delta y_i^p + \zeta_i \quad \Rightarrow \quad R_d^2 = \frac{\sigma_d^2 - \sigma_\zeta^2}{\sigma_d^2}$$

- Model-based predictions are more appropriately assessed as follows:

$$y_i^a = \alpha + \delta y_i^p + \upsilon_i \quad \Rightarrow \quad R_a^2 = \frac{\sigma_a^2 - \sigma_\upsilon^2}{\sigma_a^2}$$

An Example

- The standard method for assessing performance calculates R^2 as follows:

$$y_i^d = \alpha + \delta y_i^p + \zeta_i \quad \Rightarrow \quad R_d^2 = \frac{\sigma_d^2 - \sigma_\zeta^2}{\sigma_d^2}$$

- Model-based predictions are more appropriately assessed as follows:

$$y_i^a = \alpha + \delta y_i^p + \upsilon_i \quad \Rightarrow \quad R_a^2 = \frac{\sigma_a^2 - \sigma_\upsilon^2}{\sigma_a^2}$$

- Let $y_i^d = y_i^a + \omega_i$ and let $B = R_a^2 - R_d^2$. We can then write

$$R_d^2 = R_a^2 \left(\frac{\sigma_d^2 - \sigma_\omega^2}{\sigma_d^2} \right)$$

Data

- We use the 2015 Mexican Intercensal Survey to construct a "census" that we use to conduct simulation experiments.

Data

- We use the 2015 Mexican Intercensal Survey to construct a "census" that we use to conduct simulation experiments.
- The census consists of 3.9 million households across 1,865 municipalities and 16,297 PSUs.

Data

- We use the 2015 Mexican Intercensal Survey to construct a "census" that we use to conduct simulation experiments.
- The census consists of 3.9 million households across 1,865 municipalities and 16,297 PSUs.
- For each household, we observe income, geographic location, demographics, and dwelling characteristics, among other information.

Data

- We use the 2015 Mexican Intercensal Survey to construct a "census" that we use to conduct simulation experiments.
- The census consists of 3.9 million households across 1,865 municipalities and 16,297 PSUs.
- For each household, we observe income, geographic location, demographics, and dwelling characteristics, among other information.
- We supplement the census data with remotely-sensed data available at the municipality level:
  - Conventional GIS-type data (human-made structures, vegetation, water, etc.)
  - Night-time lights data (Earth Observation Group)

Data

- We draw 500 survey samples ($\approx$ 23,500 households each) from the census using a two-stage design:
  1. We select PSUs with probability proportional to size
  2. We select 10 households within each cluster

Data

- We draw 500 survey samples ($\approx$ 23,500 households each) from the census using a two-stage design:
  1. We select PSUs with probability proportional to size
  2. We select 10 households within each cluster
- For each sample, we then have direct poverty estimates and true poverty measures for each subnational entity.

Data

- We draw 500 survey samples ($\approx$ 23,500 households each) from the census using a two-stage design:
  1. We select PSUs with probability proportional to size
  2. We select 10 households within each cluster
- For each sample, we then have direct poverty estimates and true poverty measures for each subnational entity.
- We implement several different approaches to poverty mapping using the survey datasets:
  - Parametric versus non-parametric models
  - Census-based versus remotely-sensed covariates

Machine Learning

- We focus on one particularly popular machine-learning method: gradient boosting.

Machine Learning

- We focus on one particularly popular machine-learning method: gradient boosting.
- Originally developed by Friedman (2001), gradient boosting combines a large number of weak learners to form a stronger ensemble prediction.

Machine Learning

- We focus on one particularly popular machine-learning method: gradient boosting.
- Originally developed by Friedman (2001), gradient boosting combines a large number of weak learners to form a stronger ensemble prediction.
- Rather than averaging models, gradient boosting sequentially adds models to the ensemble by fitting new models to negative gradient of the loss function:

$$y_i^p = \sum_t f_t(x_i)$$

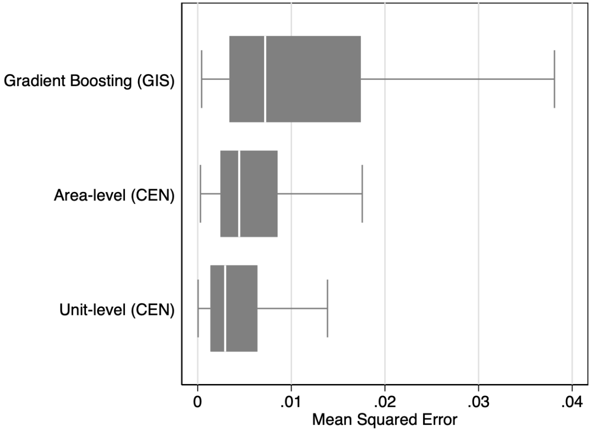
Machine Learning

- We focus on one particularly popular machine-learning method: gradient boosting.
- Originally developed by Friedman (2001), gradient boosting combines a large number of weak learners to form a stronger ensemble prediction.
- Rather than averaging models, gradient boosting sequentially adds models to the ensemble by fitting new models to negative gradient of the loss function:

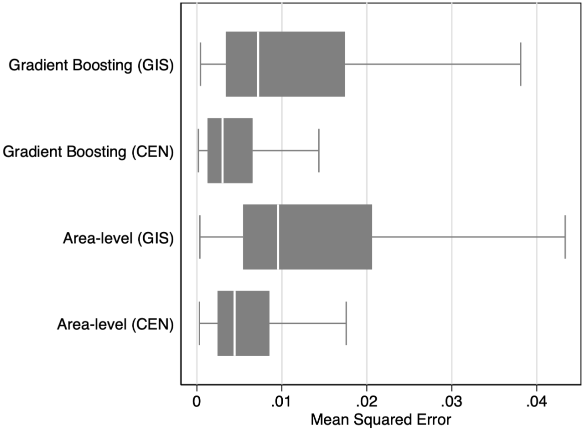
$$y_i^p = \sum_t f_t(x_i)$$

- We specifically use extreme gradient boosting (XGBoost), which uses regression and classification trees as the base models.

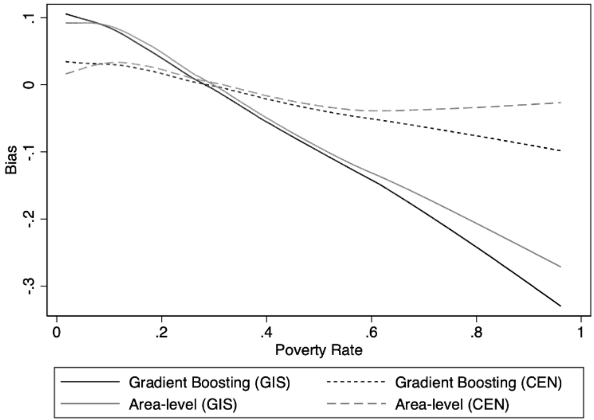
Results


Gradient Boosting (GIS)
Area-level (CEN)
Unit-level (CEN)
0
.01
.02
.03
.04
Mean Squared Error

Results


Gradient Boosting (GIS)
Gradient Boosting (CEN)
Area-level (GIS)
Area-level (CEN)
0
.01
.02
.03
.04
Mean Squared Error

Results


.1
0
-.1
-.2
-.3
Bias
0
.2
.4
.6
.8
1
Poverty Rate
Gradient Boosting (GIS)
Gradient Boosting (CEN)
Area-level (GIS)
Area-level (CEN)

Conclusions

- The results of our simulations can be summarized as follows:
  1. The modern approach to poverty mapping tends to perform poorly relative to the traditional approach.
  2. This under-performance is due to the low predictive capacity of remotely-sensed covariates.
  3. With the same covariates, machine learning is associated with more biased estimates, particularly for the poorest areas.

Conclusions

- The results of our simulations can be summarized as follows:
  1. The modern approach to poverty mapping tends to perform poorly relative to the traditional approach.
  2. This under-performance is due to the low predictive capacity of remotely-sensed covariates.
  3. With the same covariates, machine learning is associated with more biased estimates, particularly for the poorest areas.
- Main takeaway: The modern approach to poverty mapping should be used with caution due to performance limitations and risk of bias.

Conclusions

- The results of our simulations can be summarized as follows:
    1. The modern approach to poverty mapping tends to perform poorly relative to the traditional approach.
    2. This under-performance is due to the low predictive capacity of remotely-sensed covariates.
    3. With the same covariates, machine learning is associated with more biased estimates, particularly for the poorest areas.
- Main takeaway: The modern approach to poverty mapping should be used with caution due to performance limitations and risk of bias.
- Limitation: It is possible that the remotely-sensed data we use has less predictive power than the data used in other applications.