Digital Breadcrumbs & Dietary Diversity: Can mobile phone metadata predict food security outcomes among rural Haitian households?

Travis J. Lybbert (UC Davis)
with Daniel Putman, Oscar Barriga Cabanillas and Joshua Blumenstock

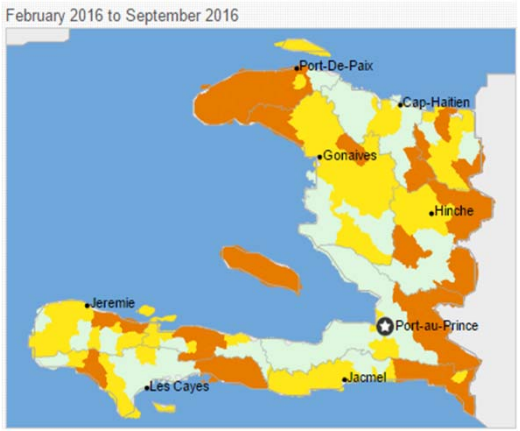
ASSA Virtual
4 January 2021

A Serendipitous & Irresistible Research Opportunity

MON Cash

Can CDR-based wealth prediction methods (e.g., Blumenstock et al., 2015) be harnessed for CDR-based impact evaluation?

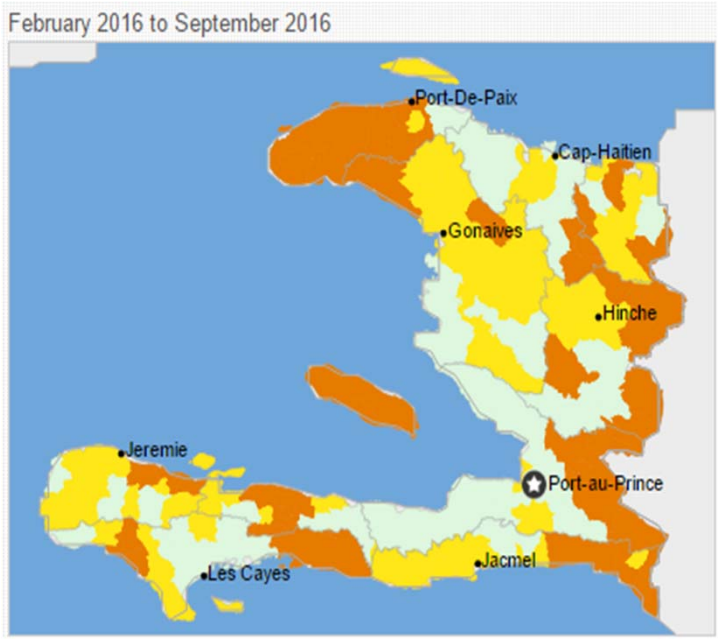
WFP World Food Programme wfp.org



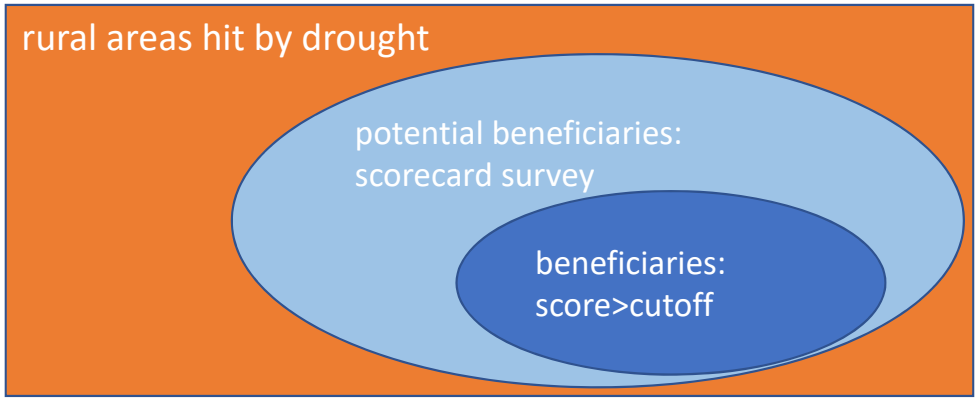
Benchmark: Conventional Regression Discontinuity (RD) evaluation of impact on food consumption using survey data

How closely can we replicate this estimated impact with a CDR-based evaluation that uses *predicted* food consumption outcomes?

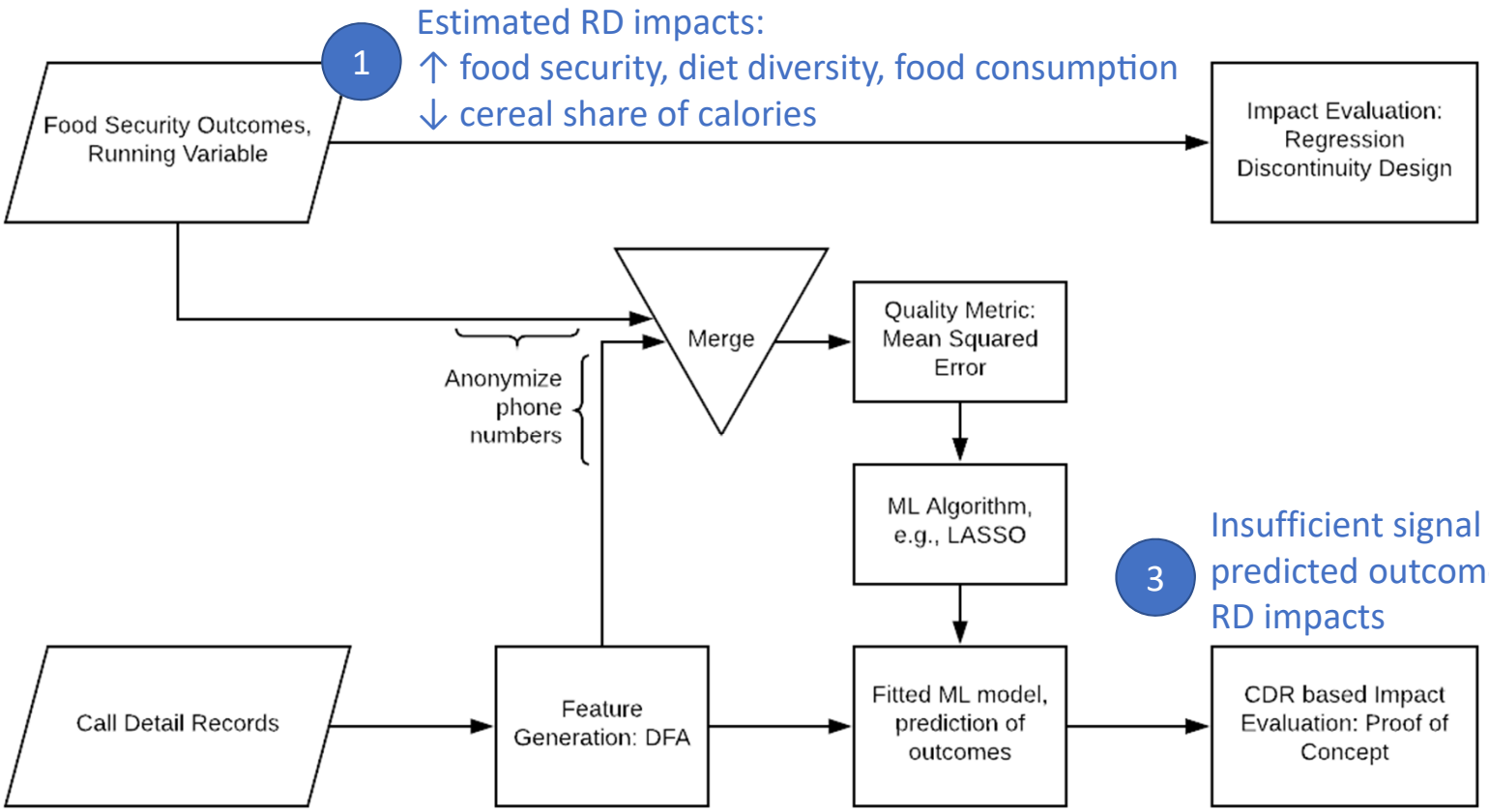
WFP's 2016 (Unconditional) Cash Transfer



- In wake of 2014-15 drought, WFP launches emergency cash transfer (EMOP) targeting the hardest hit rural areas
  - Goal: 700,000 beneficiaries (140,000 households)
  - ~$50/HH/month for 3 months
- *Mon Cash* distribution rolls out in 2016
  - New Digicel SIM cards distributed as needed
- Three tiered targeting



Research Design & Preview of Results

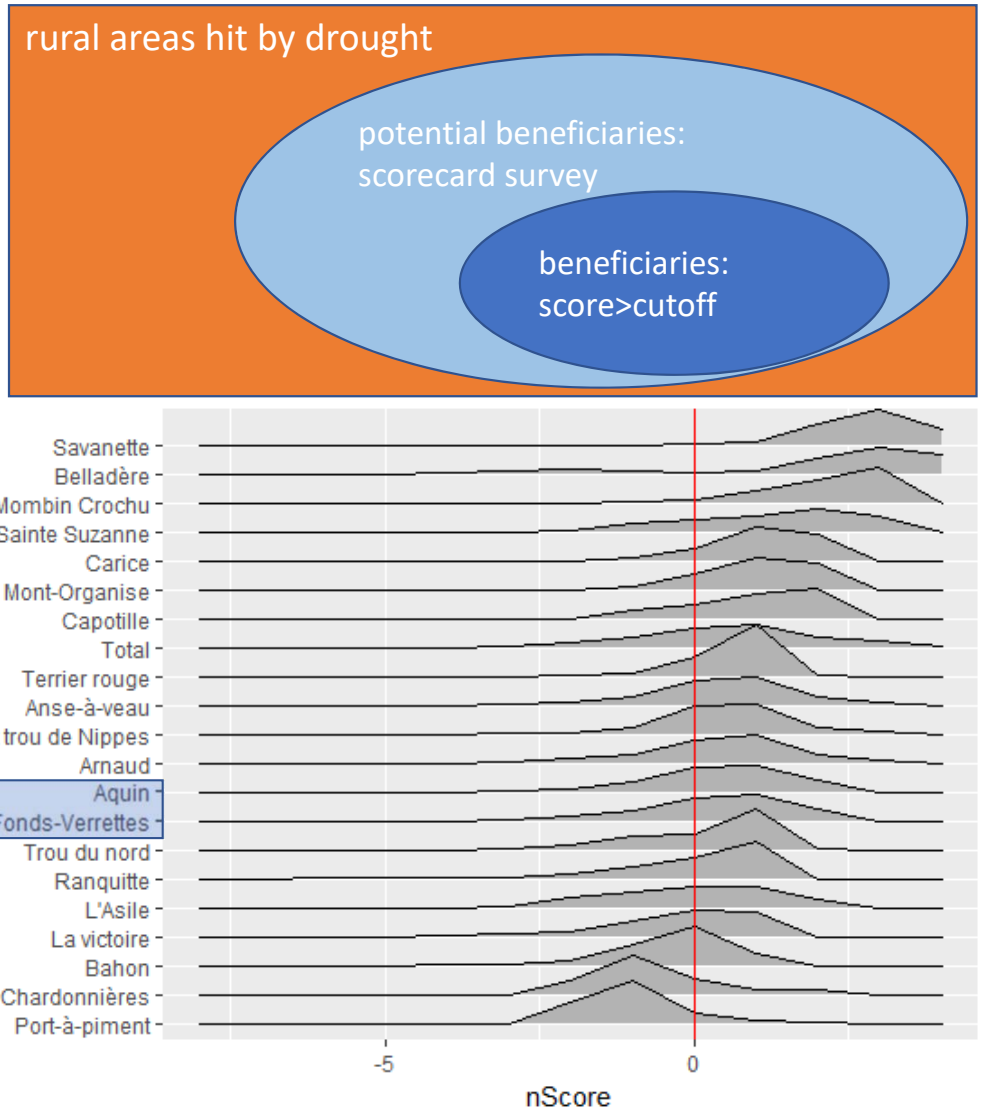


Data

1. WFP-led Scorecard
   Sampling frame: 2nd targeting tier
   Normalized score: 3rd targeting tier

2. Researcher-led HH Surveys
   Ground (in-person) survey
   Phone survey

3. CDRs from full Digicel network
   Phone numbers, time, duration, and towers connecting the calls for domestic and international calls, and SMS
   Amount and time of transfers and top-up transactions



Outcomes

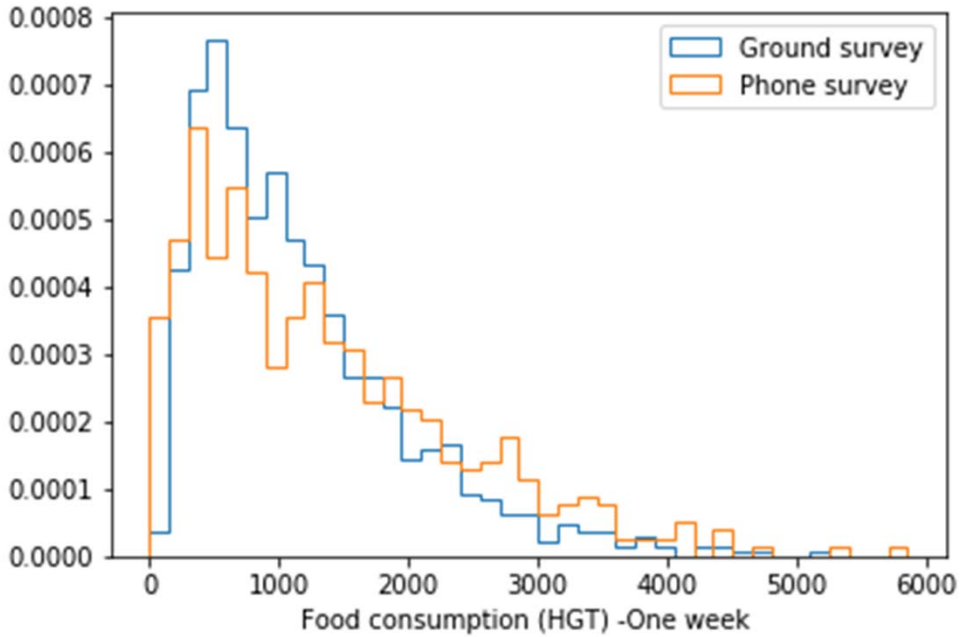
- Food Security: WFP Indices
  - Food Consumption Score (FCS)

    Number of days in past week specific food groups consumed

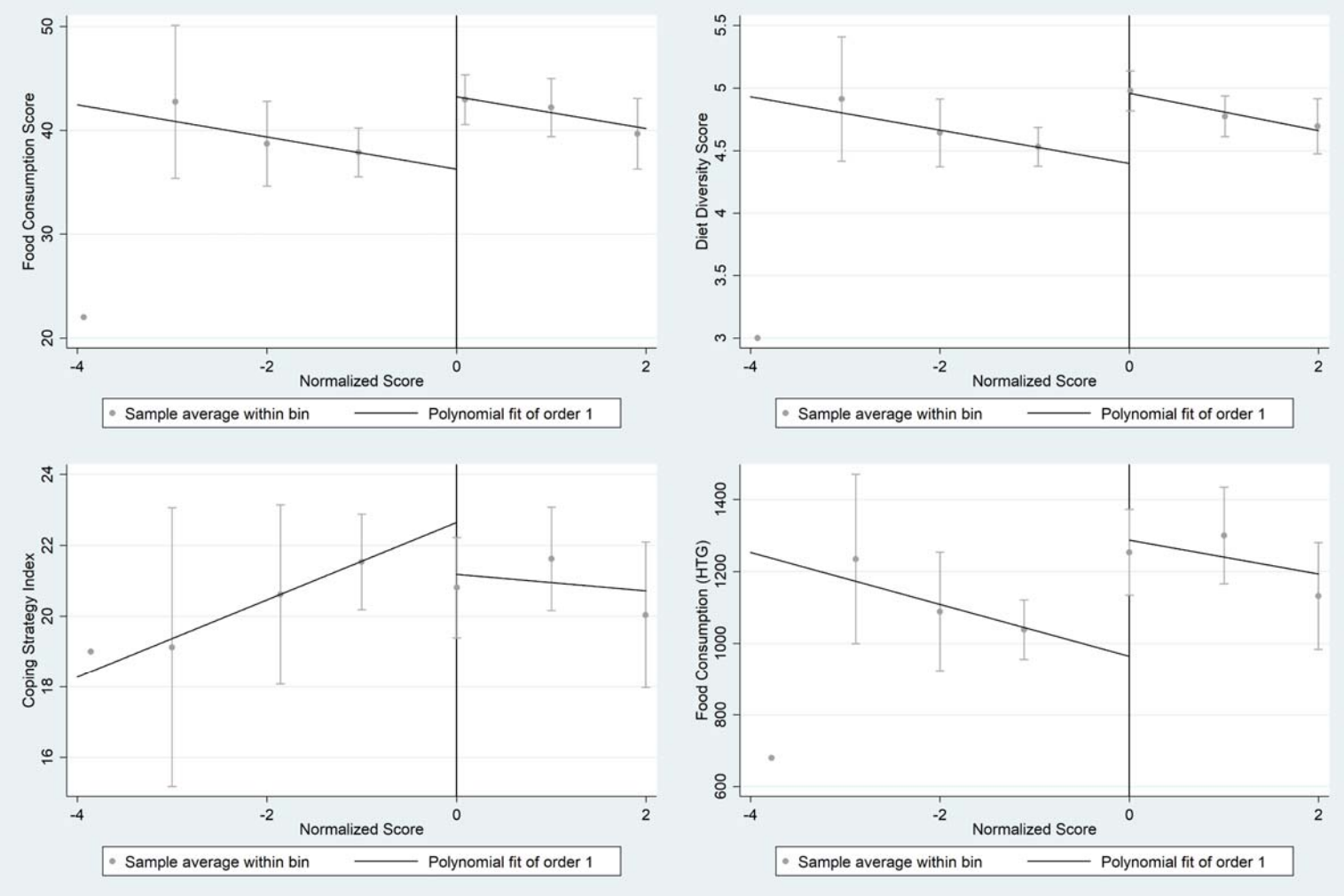
    Aggregated, weighted by nutritional value
  - Diet Diversity Score (DDS)

    Number of groups consumed over the past week
  - Coping Strategies Index (CSI)

    Number of days in past week a certain strategy used, weighted by severity
- Food Consumption



Conventional RD Results: EMOP worked



Encouraging CDR-based predictions

- Blumenstock et al. 2015 use one year of CDRs to predict HH wealth and construct a high-resolution poverty map
- Decuyper et al. 2016 aggregate CDRs and national WFP food security survey to areas w/population 10,000-50,000

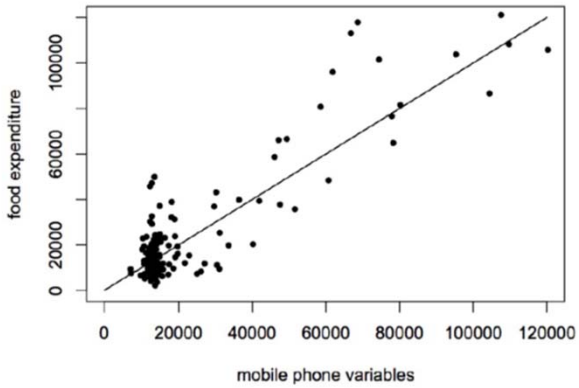
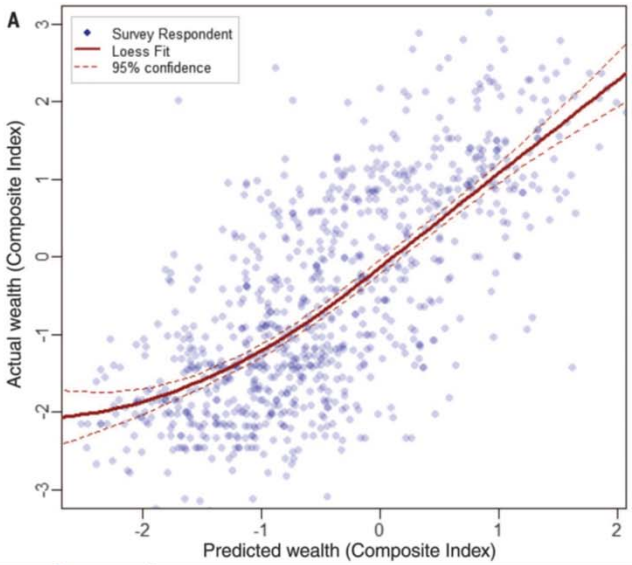
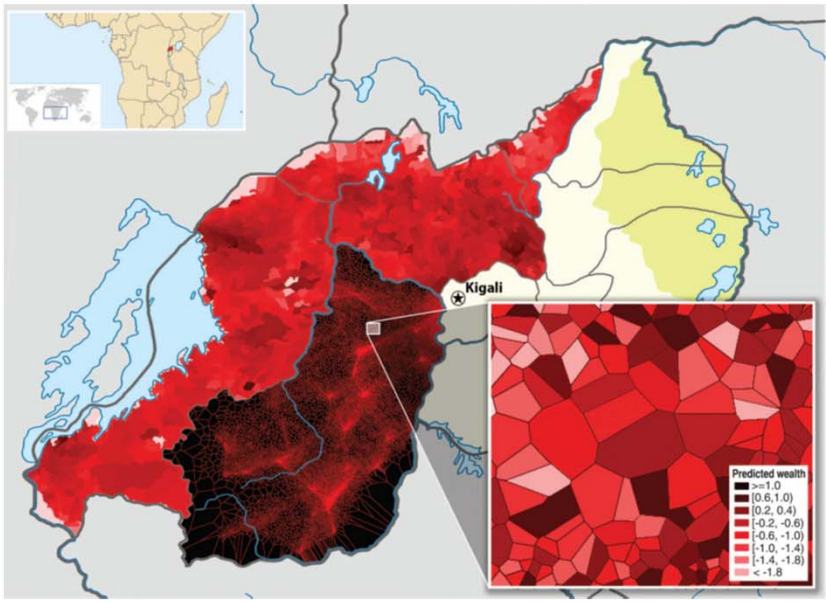


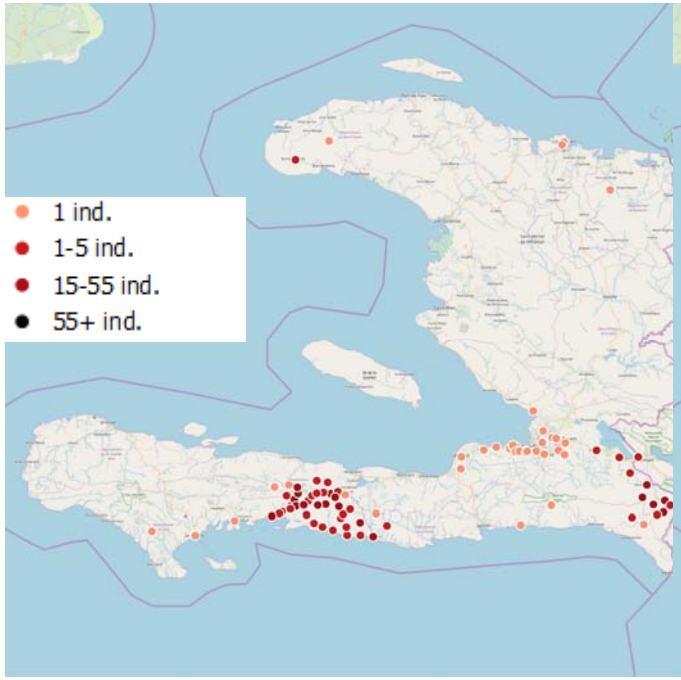
Figure 2. Quadratic combination of CDR variables against expenses on food. Correlation coefficient: 0.89.



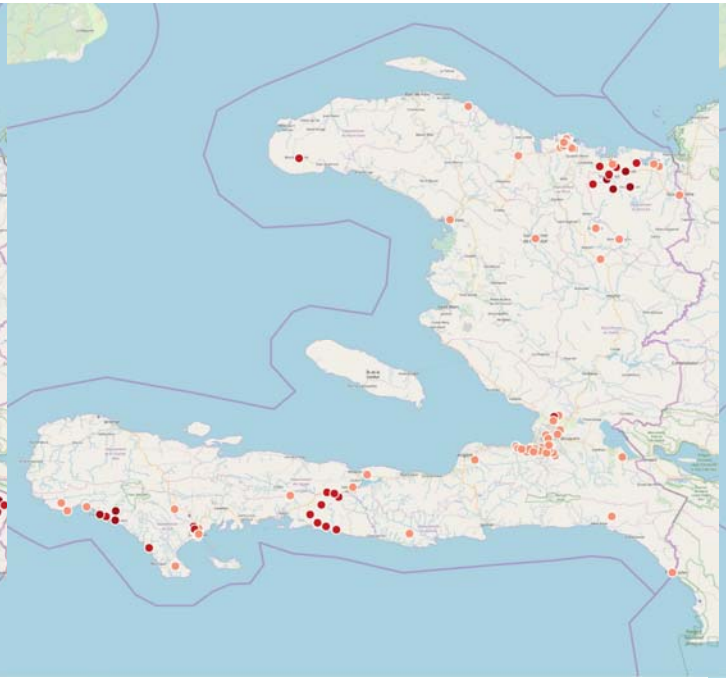


Merged Survey & CDR data:
Tower Locations of Survey Respondents

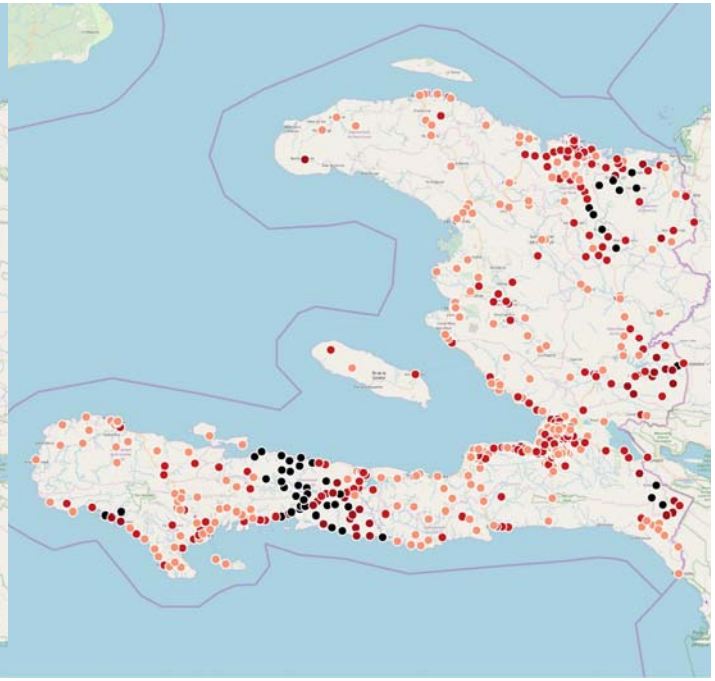
Ground survey



Phone survey

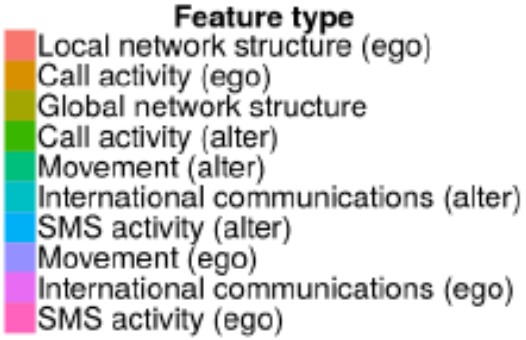


Scorecard survey

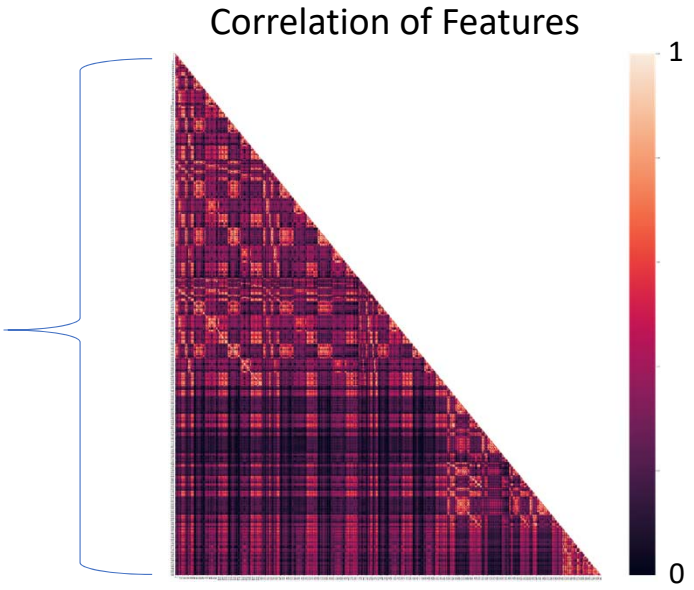


Extracting Features from CDRs

- Define Feature Automata (DFA) extracts features to predict outcomes
- The full Digicel network used to extract ‘alter’ features in this DFA
- To synchronize features with consumption outcomes, we run the DFA over 4 different (pre-survey) time-windows
- We use Bandicoot (94 features) as alternative



Time window	Ground survey	Phone survey	Features
One-year	999	550	6,447
30-days	950	550	6,447
15-days	921	537	6,220
5-days	878	526	6,109

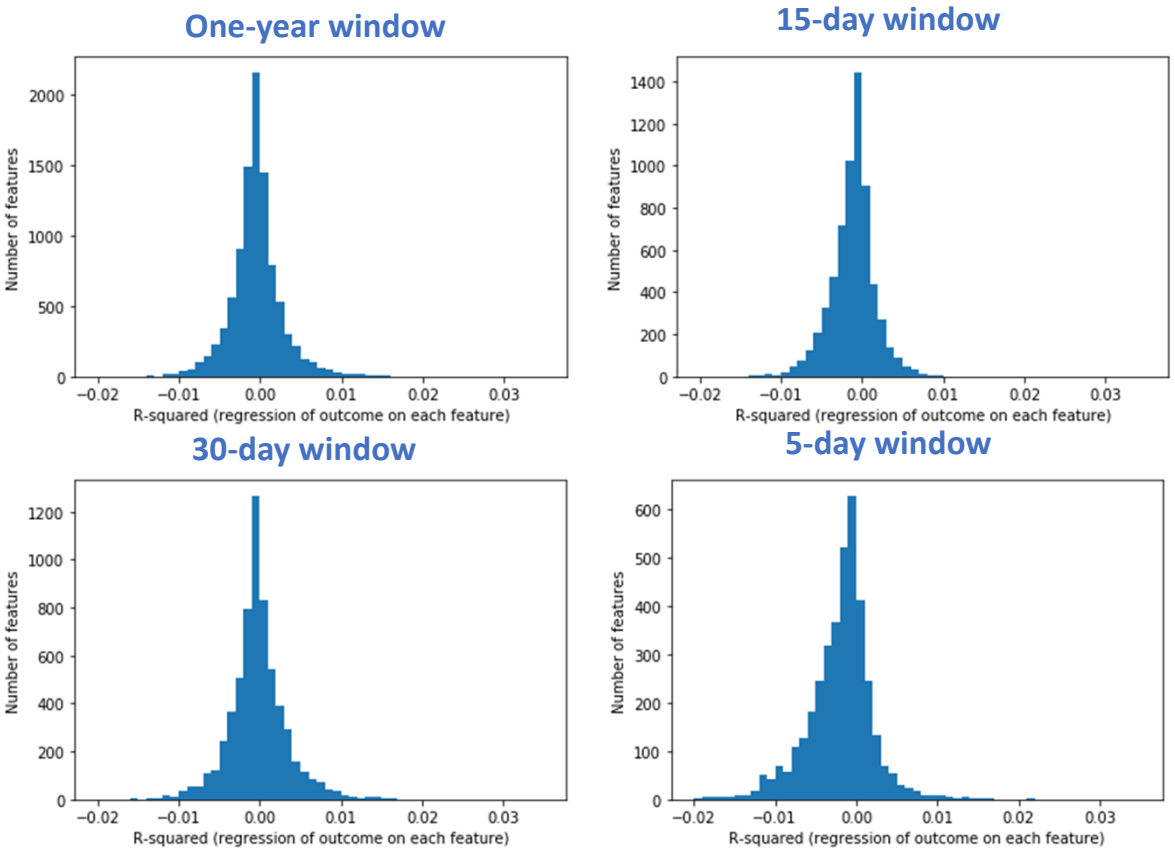


Predictive Model: ~No feature-level predictive power

Individually, our features have very little predictive power in the test data

- 5-fold cross validation
- ~No change by time window
- ~No change by survey

Predicted outcome: Food Consumption Ground survey



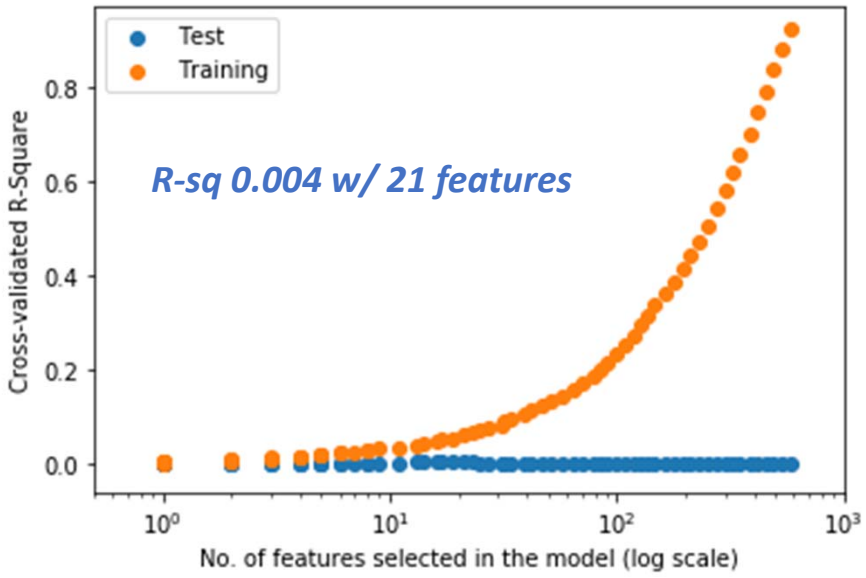
Predictive Model: Lasso doesn't do much better

- Feature selection using a 5-fold cross validation lasso model
- Slight improvement when features extracted from longer window (R-sq 0.015)
- Similarly poor prediction with alternate models
  - Ridge and elastic nets
  - Random forests
  - XGBoost

- Core problem: CDR features lack signal

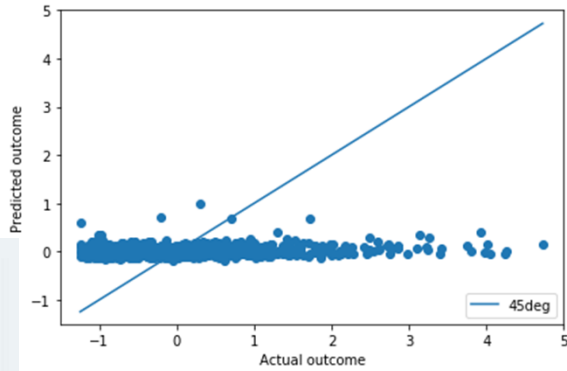
Food consumption: One week (HGT)

30-day window

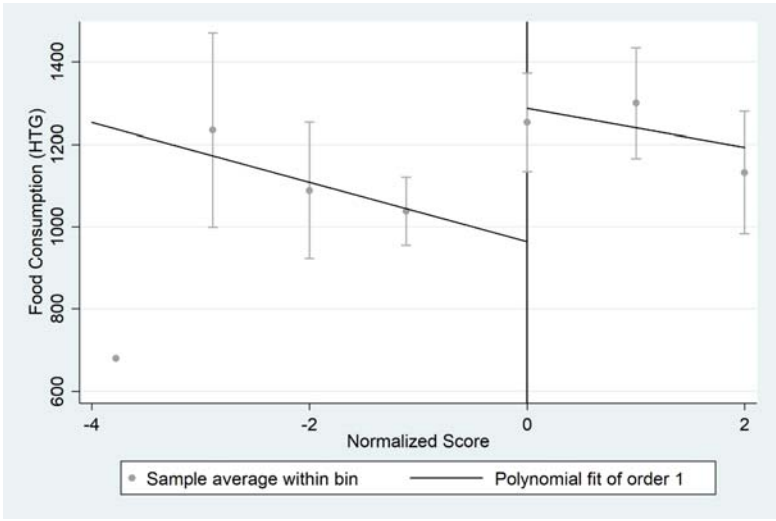


CDR-based Impact Evaluation: RD with predicted food consumption

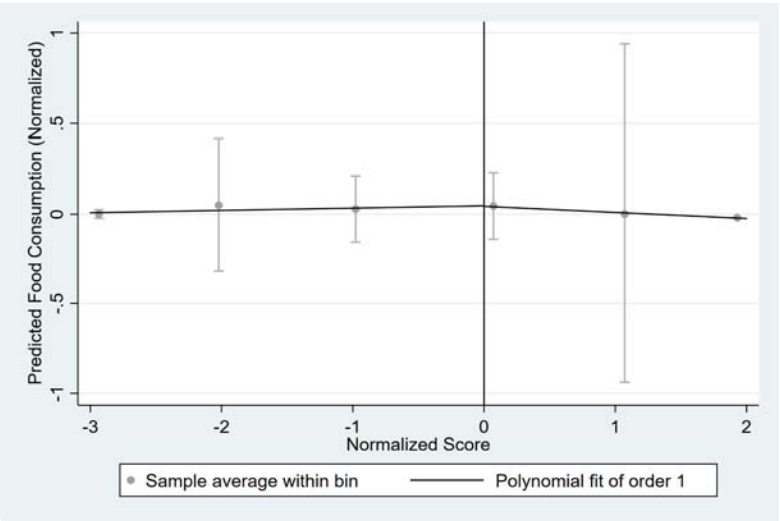
Obviously, with ~no predictive power, we can't replicate the conventional RD impact estimate



Survey Values



CDR-based Predicted Values



Post-mortum: What went wrong?

1. Known sources of noise

   CDR-related: Sharing phones or multiple phones per HH (e.g., new SIMs from EMOP)
   Hurricane Matthew (2 months prior to ground survey; ~70% of sample affected)
   Sample restrictions to avoid these known sources do not improve R-sq.

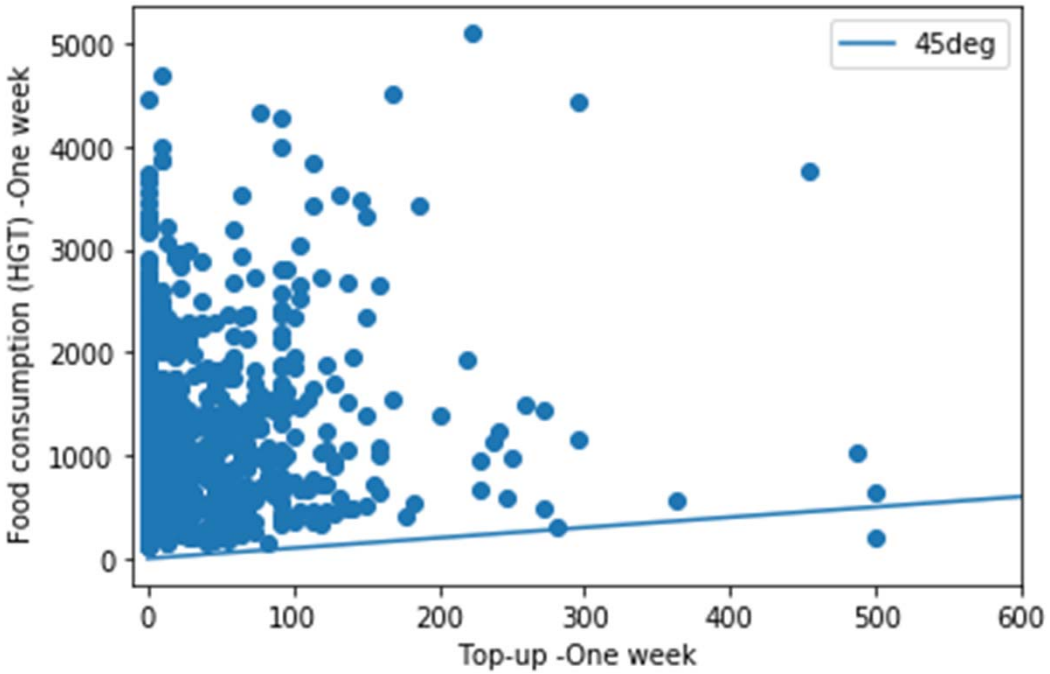
2. Limited correlation of CDRs and food consumption

   Predicting stocks (e.g., assets, wealth) easier than flows?
   Do our CDR-based features predict anything?
   Do transfers impact network activity?
   Probably. Not really. Yes.

3. Effective programmatic targeting narrows range of HH wealth

   Intentional focus on poor rural HHs limits variability of outcomes and CDRs?
   This is our primary suspect – still working on an empirical test.

2. Limited correlation of CDRs and food consumption



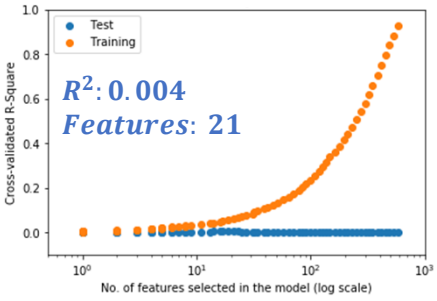
Correlation=0.12

Can our CDR features predict more stable outcomes?

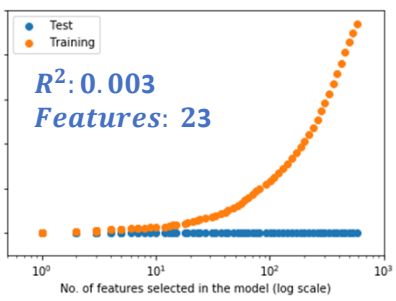
- Wealth index?
- Livestock?
- …top-up recharges?

30-day window: Ground survey

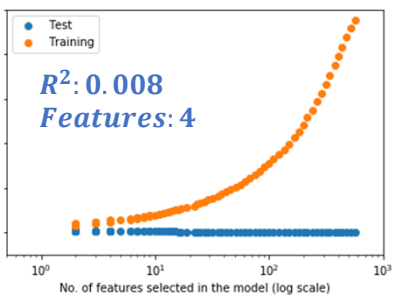
Food consumption



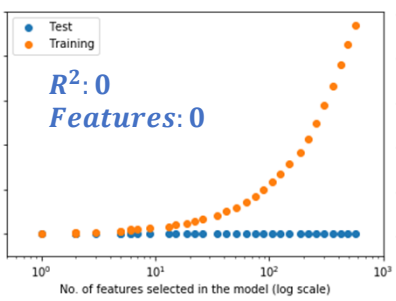
Beneficiary



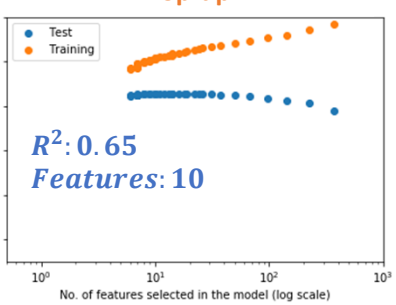
Livestock



nScore

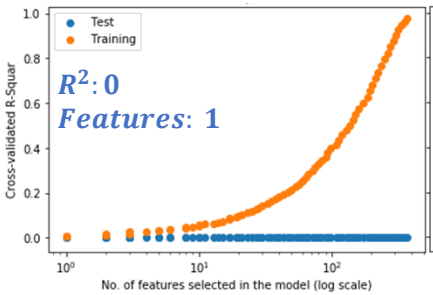


Top-up

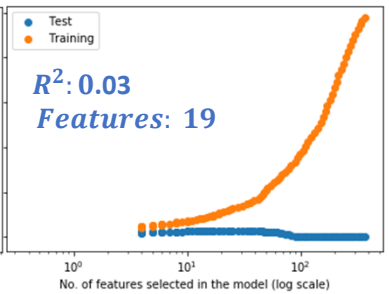


30-day window: Phone survey

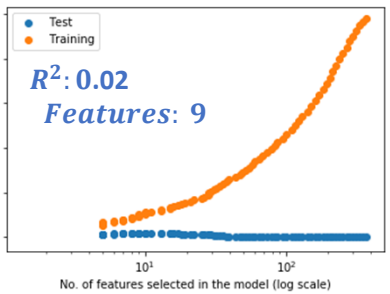
Food consumption



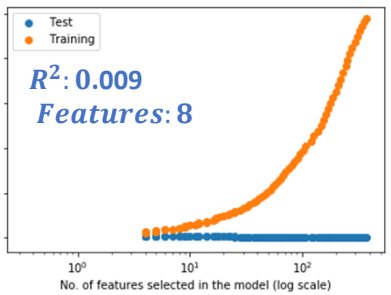
Beneficiary



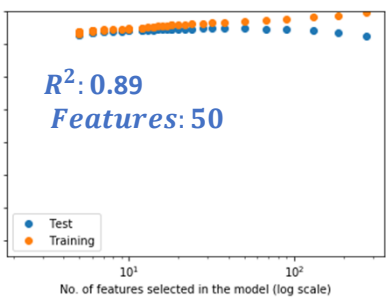
Livestock



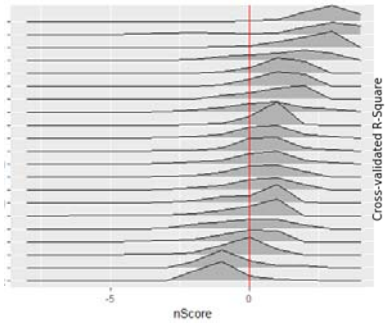
nScore



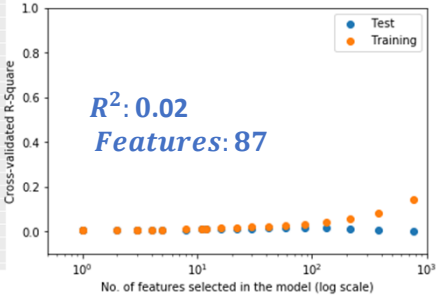
Top-up



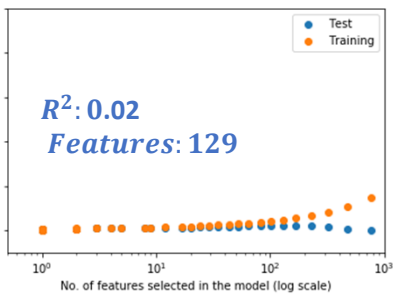
30-day window: Scorecard survey



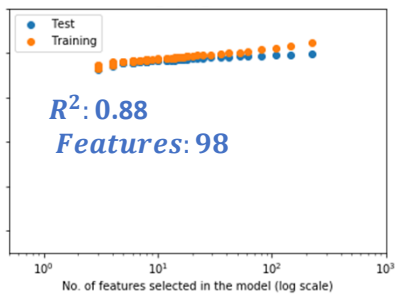
Beneficiary



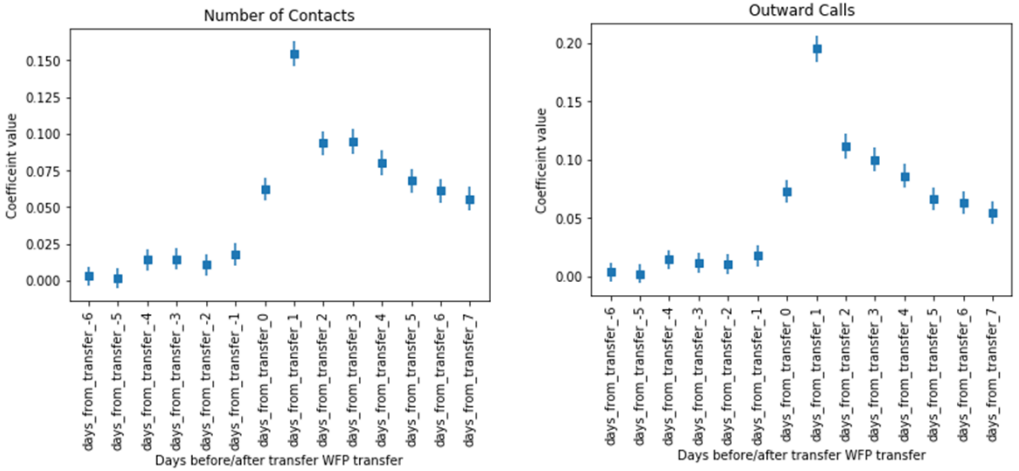
nScore

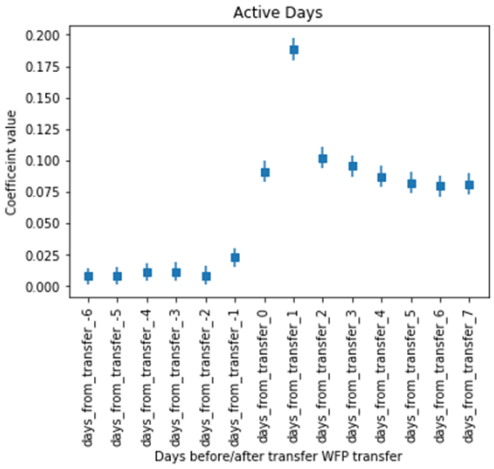


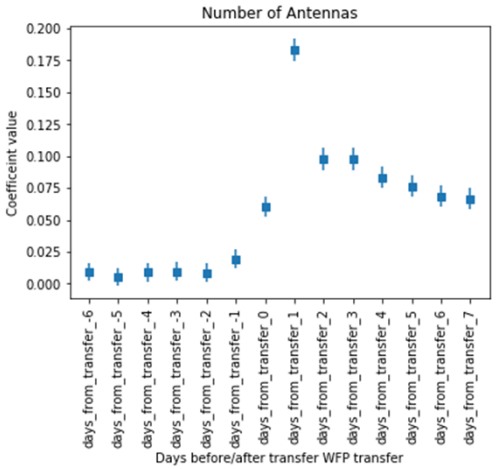
Top-up



Do EMOP transfers impact network activity at all?



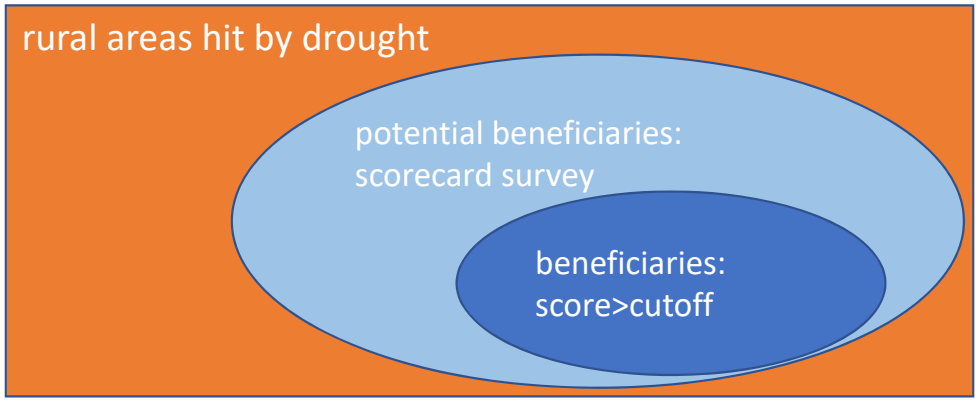


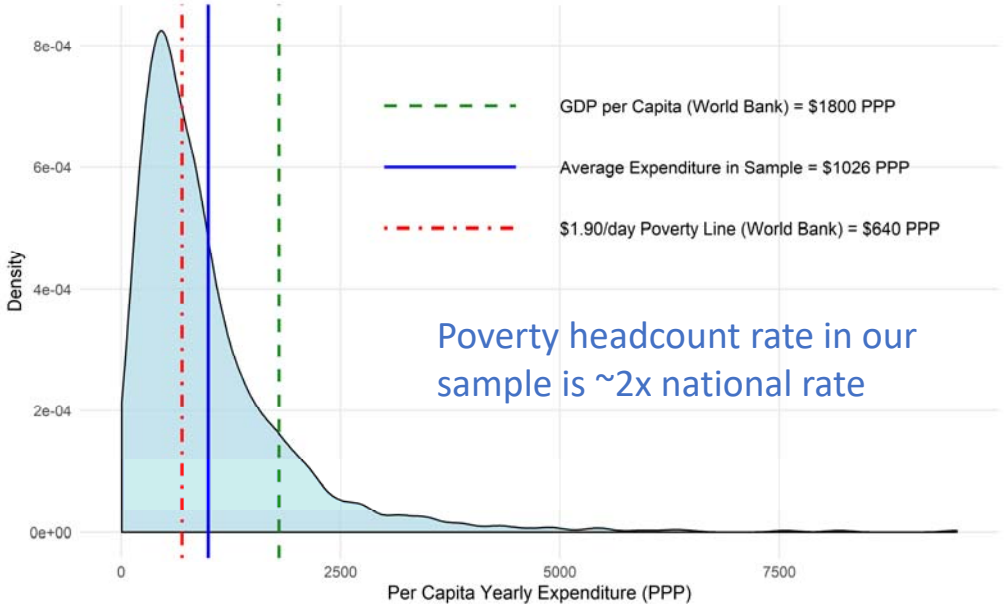


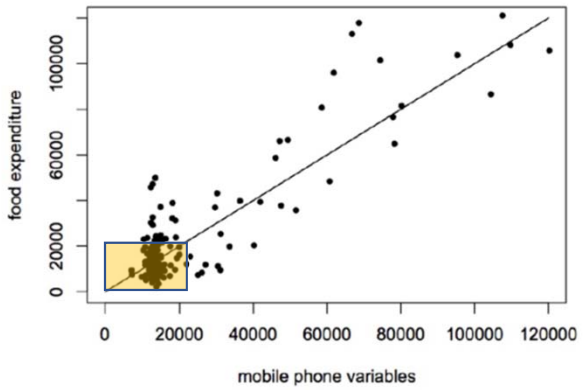
- 1 week pre/post first EMOP transfer
- N=38,657 phone numbers

- Clear spike in network activity with apparent convergence to new higher level of usage and mobility

2. Effective programmatic targeting







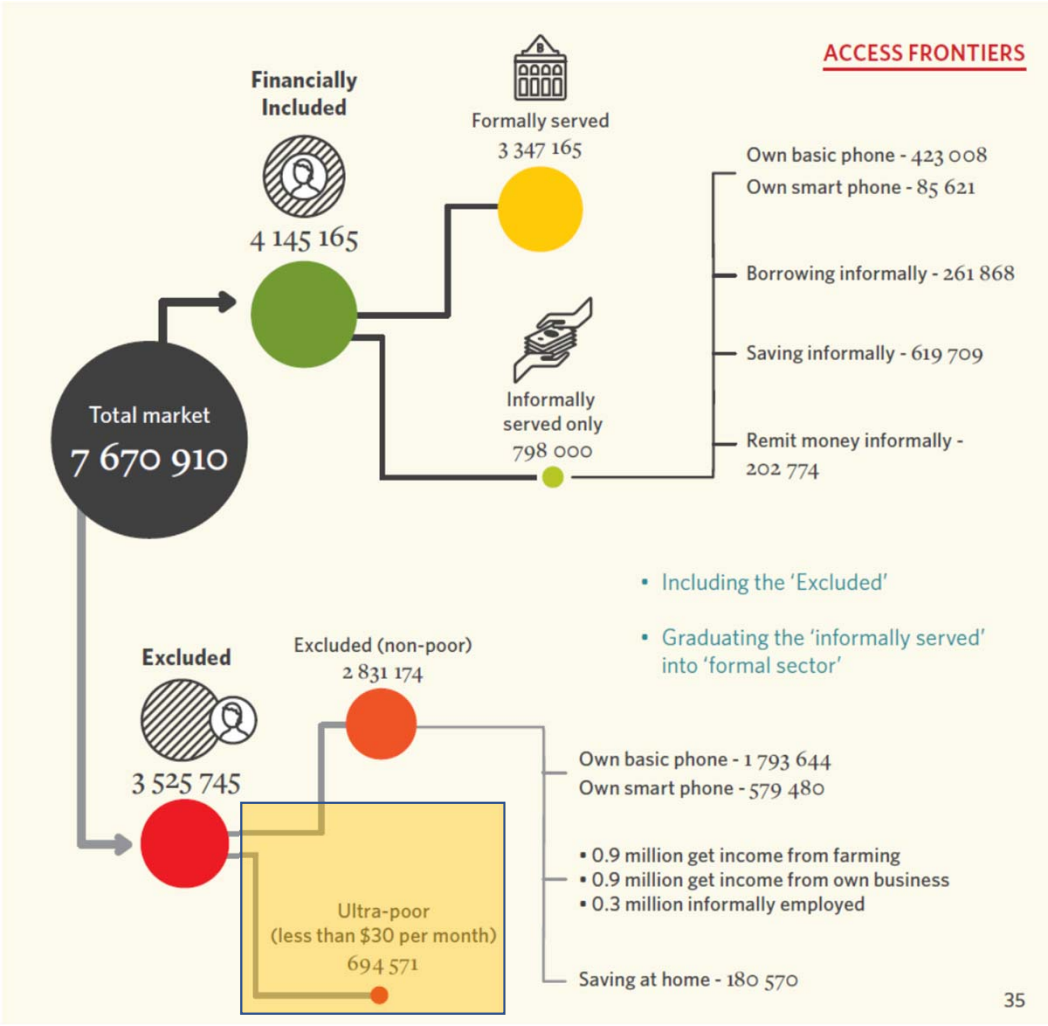
Explicit and effective targeting implied too narrow a range of outcomes and CDRs for CDR-based features to differentiate between HHs

Figure 2. Quadratic combination of CDR variables against expenses on food. Correlation coefficient: 0.89.

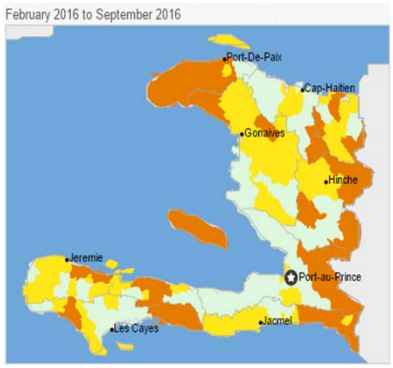
2. Effective programmatic targeting

- FinScope (USAID) 2018 data as a means of testing this suspicion
- Nationally representative (N=4,269)
- Detailed in-person survey
- ~Half are included in our CDR data
- Follow up survey with expanded informed consent underway
- Will allow us to test predictive power of CDR-based features across nationally-representative sample
- How much predictive power do we lose as we narrowly focus on the poor?

Source: FinScope Report (2019)



Final Thoughts



Can CDR-based wealth prediction methods be harnessed for CDR-based impact evaluation?

- Not in this context, which seemed ideal in some ways
- If targeting undermines this test, then perhaps there is fundamental tension lurking between these methods
- Impact evaluation hinges on good counterfactuals
- Sharp focus on treated and comparable counterfactual units helps identification but hurts CDR-based prediction

- Better suited for interventions with coarser units of E(impact)?
  - Investments in irrigation or transportation infrastructure
  - Changes in local institutions or incentives
  - Local economy spillovers
- Lingering (untested) limitations on temporal resolution of CDR-based prediction?