www.sciencemag.org/content/350/6264/1073/suppl/DC1

Supplementary Materials for

Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,* Gabriel Cadamuro, Robert On

*Corresponding author. E-mail: joshblum@uw.edu

Published 27 November 2015, *Science* **350**, 1073 (2015)
DOI: 10.1126/science.aac4420

This PDF file includes:

MATERIALS AND METHODS

I. Data Description and Construction

A. Phone survey administration

In Summer 2009, we coordinated a phone survey of a geographically stratified group of Rwandan mobile phone users. Using a trained group of enumerators from the Kigali Institute of Science and Technology (KIST), a short, structured interview was administered to roughly 900 active mobile phone subscribers. The survey instrument contained approximately 80 questions that focused on basic socioeconomic and demographic information, including asset ownership and household and housing characteristics (Table 1). Several of these questions were drawn directly from the survey instruments used by the National Institute of Statistics of Rwanda in their Demographic and Household Surveys (DHS), which is described in greater detail below. Aside from the phone number of the respondent, we did not solicit any personally identifying information such as first name, last name, or address.

Full details on the administration of this phone survey are discussed in (*30*). In brief, the survey population was intended to be a representative sample of active subscribers on Rwanda's largest mobile phone network. At the time, the operator had roughly 90 percent market share, and 1.5 million registered Subscriber Identification Modules (SIM cards). However, since the number of registered SIMs greatly exceeds the number of active subscribers, we eliminated numbers which had not been used at least once in each of the three most recent months for which mobile phone data was available (October through December 2008). Each of the remaining 800,000 numbers was assigned to a geographic district based on the location of the phone for the majority of calls made (see SM Section IVA for details). The final sample was a geographically stratified random set of these numbers, with sampling weights determined by the distribution of active subscribers across districts (*30*).

Enumerators made three attempts to contact each respondent, on different days and at different times of day. Respondents were compensated RWF500 (roughly US$1) for participating in the survey, which took between 10 and 20 minutes to administer. Survey enumerators requested informed consent from each respondent, in which the goals of the study were described and oral permission was received to merge survey responses with anonymized call records, in accordance with the protocols of our university's ethical review board.

The contact rate was roughly 61%; non-contacts were largely the result of phones that were turned off or disconnected. The cooperation rate was 97%; almost everyone who picked up the phone was enthusiastic to participate in a study with university researchers, with whom they generally had little prior contact. We thus interpret the survey sample as representative of the population of active mobile phone subscribers, who we assume are systematically different from both the population of mobile phone subscribers and the general Rwandan population. In SM Section V, we discuss in greater detail the extent to which the non-representativeness of the phone survey sample affects our results.

B. Mobile phone call detail records (CDR)

From Rwanda's near-monopoly mobile phone operator, we obtained a complete historical log of call detail records (CDR), which contain basic metadata on all transactions mediated by the mobile phone network. The logs included all domestic and international calls, as well as every text message (SMS) sent and received on the network, from early 2005 to mid-2009. For each of these transactions, we observe the time and date of the call, the anonymized but unique identifier of the calling and receiving party, the duration of the call, as well as the cellular towers through which the call was routed. As described in greater detail in SM Section IVA, information on these cellular towers can be used to infer the approximate location of both the caller and the receiver at the time of the call. For the sample of phone survey respondents who completed the survey, information from the mobile operator was provided to match the true phone number to the anonymized identifier in the CDR dataset.

C. Demographic and Health Surveys (DHS)

To provide further validation of the external validity of this method, we compare out-of-sample wealth predictions to "ground truth" Demographic and Health Surveys (DHS) collected by the National Institute of Statistics of Rwanda. Two rounds of these surveys are used in our analysis: DHSV, which was conducted between December 2007 and April 2008 on a sample of 7,377 households; and DHSVI, conducted between September 2010 and March 2011 on a sample of 12,792 households.

The DHS surveys are conducted with a nationally representative sample of households. Villages were selected with probability proportional to village size, and households are given survey weights to allow for reconstruction of nationally representative statistics (*31*). DHSV contained 247 village clusters, while DHSVI contained 492. The geographic coordinates of each cluster's centroid are also provided with the DHS data. However, as noted in the DHS documentation, "the data are randomly displaced up to 5 kilometres in rural areas and up to 2 kilometres in urban areas. A further 1 percent of rural clusters are displaced up to 10 kilometres." These displacements add considerable measurement error to subregional estimates of wealth, but should not estimates aggregated at the district level, as is the case in most of our analysis.

D. Composite wealth index construction

In Rwanda, as in most developing countries, it is difficult to estimate the socioeconomic status of a survey respondent with a single survey question. Instead, household surveys typically rely on a large number of questions which can be used to infer the consumption or permanent income of the respondent (*32*). The Rwandan DHS, for instance, contains roughly seventy questions related to household assets, characteristics, and expenditures. The first principal component of these responses is commonly treated as a proxy indicator of the respondent's unobserved wealth (*21*).

In our phone surveys, which were designed to be very short, we did not have the option of asking such a large number of questions related to assets and housing characteristics. Instead, we selected the subset of questions that, in the DHS data, were most highly correlated with the first principal component of the full set of DHS responses. We further excluded questions that would be difficult to administer in a phone survey (e.g., in our piloting we found that most respondents were unable to quickly ascertain how much land they owned). The final set of asset-related questions is listed in Table S1B. We also include the size of the household and the number of children, but all results are robust to the exclusion of these factors.

We compute the "composite wealth index" as the first principal component of the asset and household characteristics questions in our phone survey (*21*). The basis vectors W of the covariance matrix are estimated using weighted principal component analysis on the normalized data from the 856 phone survey respondents, where the weights are determined as described in SM Section 1A above (*33*). The first principal component captures 26 percent of the total

variation in assets and household characteristics. When we later validate the phone-based predictions against data collected through government surveys (SM Section IV), we use the same basis vectors $\boldsymbol{W}$ computed on the phone survey data to project each DHS household's asset responses onto an analogous composite wealth index.

E. Satellite data

We validate the phone-based predictions of regional electrification using data on satellite "night lights" using average radiance composite images from the Visible Infrared Imaging Radiometer Suite Day/Night Band (VIIRS-DNB). The VIIRS-DNB imagery recognizes wavelengths from green to near-infrared, and is preprocessed by the National Oceanic and Atmospheric Administration to remove stray light and emphasize light from cities. The satellite data is provided at the resolution of 0.742km x 0.742km grid cells, and is measured in units of nanowatts/cm^2/square radian (*34*).[1]

II. Feature engineering

Our goal in engineering features is to transform an individual's mobile phone transaction logs into a set of quantitative metrics that in turn can be used to infer that same individual's economic state. In the related literature, the most common approach has been to carefully construct a small number of intuitive indicators from the phone metrics, and compare regional aggregates of those phone metrics to regional socioeconomic indicators. In such work, for instance, there is evidence that the geographic diversity and reciprocal nature of social relationships are both correlated with economic outcomes (*8*, *35–38*).

Our goal is different. We seek to develop measures of poverty and wealth that maximize predictive accuracy, possibly at the expense of the interpretability of the model. Thus, instead of devising a parsimonious set of metrics based on intuition, we take a brute force to feature engineering that is designed to capture as much variation as possible from the raw call detail records. Specifically, we develop a method based on a deterministic finite automaton (DFA) (*39*) to generate a large number of potentially correlated metrics, and then rely on regularization and

[1] April 2012 version. These data were obtained from the NOAA National Geophysical Data Center, Earth Observation Group.

related techniques to eliminate redundant metrics from the model. The primary advantage of using the DFA is that is restricts the number of degrees of freedom in the hands of the researcher; rather than specifying hundreds or thousands of "features" one by one, the DFA allows the researcher to specify a small number of different operations, which are then recursively applied to generate a large number of features.

A. Baseline models: Single feature and top-5 features

In addition to the combinatoric deterministic finite automaton (DFA) described below, we implement two simple approaches to establish baselines for comparison. The first is a an "intuitive" model, which consists of five hand-picked features based loosely on related work (*8*, *35–38*), and which are chosen to capture a variety of the behaviors reflected in mobile phone transaction logs. These features are: (i) the total number of calls in which the individual is involved (outgoing + incoming); (ii) the total number of text messages; (iii) the total number of international calls; (iv) the degree centrality of the individual (i.e., the total number of unique contacts with whom the individual interacts); and (v) the Radius of Gyration, a measure of the average travel distance of the individual (*15*). A cross-validated ordinary least squares model using these five features explains 20% of the variation in composite wealth index in the sample of 856 survey respondents (Table S1A).

The second baseline uses the single feature, generated by the DFA, which is empirically determined to be most strongly correlated with the wealth composite index in the sample of 856 survey respondents. More precisely, for each of the 5,088 features generated by the DFA, we use 5-fold cross-validation to divide the set of 856 respondents into five different training and testing sets (with an 80%-20% split). For each training set, we fit a linear regression of the response variable on the single feature, and compute the R^2 for the corresponding test set; we average the test R^2 across these five folds, and select the feature that has the highest average test R^2. The single best predictor, which has an average test R^2 of 0.39, indicates, for an individual *i*, the weighted average of all of *i*'s first-degree neighbors "day of week entropy" of outgoing SMS volume, where the weights are determined by the frequency of interaction between *i* and the neighbor. Roughly, this is an indication of the extent to which there is a high degree of predictability in the days of the week on which *i*'s friends and family tend to send text messages.

While this single feature performs surprisingly well, we do not expect that this could have been foreseen in advance or that it would be as informative in other contexts (see SM Section VB).

B. Deterministic finite automaton (DFA)

Our deterministic finite automaton takes as input a list of call detail records (CDRs), where each element in the list is a transaction record containing a tuple of fields (date, time, userID, and so forth). From this initial state, the data transitions to subsequent states, where each transition defines a legal operation that transforms the data input to the state into a different dataset output from the state. The final output from the DFA is a single numerical value, which is equivalent to a single behavioral metric, or "feature." Thus, any feature used in our analysis can be generated by a complete traversal of the automata.

The DFA used for feature generation is shown in SM Figure 1, and is defined by:

- A set of states $Q = \{q_0, q_1, q_2, q_3, q_{11}, q_{12}, q_{13}, q_{21}, q_{22}, q_{23}\}$
- The start state q_0
- The accepting state q_3
- The alphabet $\Sigma = \{CDRs, Fields, Field, Value\}$
- The transition function $\delta: Q \times \Sigma \rightarrow Q$

SM Figure 1 depicts the transition function. Note that we assume that any element of $Q \times \Sigma$ not pictured results in a transition from state q_i back to q_i. For example, $\delta(q_{13}, a') = q_3$ while $\delta(q_{13}, f) = q_{13}$ since it is not legal. The transition function is specified as:

- $f(\cdot): Q \times CDRs \rightarrow CDRs$: "Filter" operations on a set of CDRs that select a subset of the CDR tuples.
  Example: Filter all rows that are not incoming calls.
  Legal transitions: calls over 60 seconds; calls made during the working week (Monday - Friday, 9am -5pm); calls not made during the working week; incoming activity; outgoing activity; international activity; text messages (SMS).
- $m(\cdot), m'(\cdot), m''(\cdot): Q \times CDRs \rightarrow list(CDRs)$: "Group By" operations that transform a dataset of type D into a map from the attribute to subsets of D, where subsets are defined

by the attribute, which may be the identity of the subscriber ("ego"), the identity of the subscriber's contacts ("alter"), or a time period attribute.
Example: Group all CDRs by ego and week of the year.
Legal transitions: group by ego; group by alter; group by week.

- $s(\cdot): Q \times CDRs \rightarrow Fields$: "Select" operations that transform a set of rows into a set of values. This can be any operation on a set of rows that maps each row to a number.
Example: Select a single field from a row (such as "duration of call")
Legal transitions: select duration of event; select geocoordinates of ego at time of event; select day of week; select hour of day;

- $a(\cdot), a'(\cdot), a''(\cdot): Q \times Field \rightarrow Value$: "Aggregate" operations that aggregate a set of numbers into a single number. These convert a mapping from some attribute to a set of values to a mapping from an attribute to a single number.
Example a: Compute mean of a list of numbers; compute radius of gyration of a set of geocoordinates.
Example a': Computes aggregation over first-degree network properties, e.g., average PageRank of first degree neighbors of an individual.
Example a'': Computes aggregation over time, e.g., trend over time in calls per week.
Legal transitions: mean; maximum; minimum; standard deviation; sum; radius of gyration; count of unique values.

- $r'(\cdot), r''(\cdot): Q \times CDRs \rightarrow CDRs$: "Reduce" operations that groups a multi-level mapping by one level.
Example: Subsets grouped by user-time are aggregated into subsets grouped by user.
Legal transitions: Values mapping to (ego, time period) tuples are grouped in sets identified by the ego and the mapping from egos to their sets are returned.

Not depicted in the DFA, but also included in feature engineering, are simple transformations (log and quadratic) of the DFA traversals. We also experimented with including features that do not fall neatly into this framework, such as PageRank, but in practice this has little effect on the results.

As an example, the following traversal of the DFA will produce a feature that indicates the standard deviation of the weekly average call duration during working hours.

- Start state q_0
- $\delta\left(q_0, f_{workday} \in f\right) = q_0$: filters out calls made on weekends or outside 9am-5pm
- $\delta(q_0, m) = q_1$: groups all calls by subscriber ("ego")
- $\delta(q_1, m'') = q_{21}$: groups all calls by week, calls are now grouped by subscriber-week
- $\delta(q_{21}, s_{duration} \in s) = q_{22}$: converts groups of calls to groups of call-durations
- $\delta(q_{22}, a_{mean} \in a) = q_{23}$: computes the mean of each group of call-durations; at this point, each subscriber is represented by a set of weekly averages
- $\delta(q_{23}, a''_{sd} \in a) = q_3$: computes the standard deviation of weekly averages.

C. Feature categorization

While the DFA is effective in constructing a large number of features from a relatively parsimonious grammar, the quantity of resultant features complicates interpretation. This is a clear disadvantage of the DFA relative to more parsimonious models based on intuitive features. As noted above, however, our primary goal is predictive accuracy, not model interpretation. Nonetheless, to inform the subsequent analysis, we label each feature with a "type" by grouping the features according to approximate function. Alternative partitionings of the feature space are equally plausible, but the partition we choose roughly follows the broad classes of features discussed in related literature (*40*).

- SMS activity (ego): Metrics reflecting SMS-based activity of the subscriber, including volume, variance, variation over time, etc.
- SMS activity (alter): Metrics reflecting SMS-based activity of the subscriber's first-degree network (FDN).
- Call activity (ego): Metrics of call-based activity of the subscriber.
- Call activity (alter): Metrics of call activity of the subscriber's FDN.
- International communications (ego): International call activity of the subscriber.
- International communications (alter): International call of the subscriber's FDN.
- Movement (ego): Information on the pattern of locations visited by the subscriber

- Movement (alter): Information on the locations visited by the subscriber's FDN.
- Local network structure (ego): Simple properties describing the subscriber's position within his or her FDN.
- Local network structure (alter): properties of the subscriber's FDN's social networks.
- Global network structure: Structural properties describer the subscriber's position within the entire graph, such as PageRank and clustering coefficients.

III. Model fitting and out of sample prediction

A. Supervised learning

From the several thousand behavioral metrics constructed by the DFA, we used supervised learning techniques to identify a smaller subset of features that are the best joint predictors of the response variable, using the sample of 856 survey respondents to train the model. Specifically, we use elastic net regularization (*41*) to penalize model complexity and reduce the likelihood that the model is "overfit" on the small number of training instances. For each possible model parameter β_j, the elastic net imposes a penalty equal to

$$\lambda \sum_{j=1}^{p} \left(\alpha \beta_j^2 + (1-\alpha) \left| \beta_j \right| \right), \quad (1)$$

This penalty linearly combines a lasso (L_1) penalty for variable selection with variable shrinkage as in ridge regression (L_2), where higher values of λ produce more parsimonious models. As noted in SM Section IIA, we compare the elastic net model to models using lasso and ridge regression separately, and find only modest differences in performance from the elastic net. Similar results obtain when using nonlinear tree-based ensemble regressors to predict the continuous-valued composite wealth index, and random forest classifiers to predict asset ownership and housing characteristics (*24*) – these results are presented in Table S1A.

For each model, we use cross-validation to help ensure that the model will generalize beyond the small sample upon which it is fit. Specifically, we use 5-fold cross-validation to select model parameters that maximize average R^2 on the held-out test data across 5-folds.[2] Each fold is

[2] Cross-validation is a common method for model selection and validation. The data is first randomly divided into K random subsets, called "folds." Then, each fold is removed from the dataset, one at a time; the model is fit on the remaining data, and evaluated on the held-out fold. This process is repeated for each fold, and the model

selected with a weighted bootstrap, where the weights are determined as described in SM Section SIA to help ensure that the model is representative of the total population of mobile phone subscribers (*43*).[3]

SM Figure 4A illustrates how model performance depends on the choice of the regularization parameter λ. For large values of λ, the model selects a very small number of features, and the average performance on both the training and testing data is quite poor. (For extreme values of λ, performance is also considerably worse than the unregularized single-predictor model). As λ is decreased, a larger number of features enter the model, and performance on both the training and testing data increases until the optimal model selects 101 features. Additional increases in λ yield improved performance on the training data, but performance on the test data degrades as the model is overfit to the training instances.

B. Improving model performance

While model performance appears to be only marginally affected by the choice of the learning algorithm, we find that predictive performance is significantly impacted by the relatively small number of independent observations available. This issue is illustrated in SM Figure 2, where we show the performance that would have been achieved if we had trained on a smaller number of independent observations. These hypothetical scenarios are determined by drawing a random subset of m observations from the full set of 856 respondents, then re-training the model as if only those observations were available. We interpret the monotonic increase with sample size, and the continued positive slope at the maximum where m=856, as evidence that further performance gains could be achieved by expanding the sample of phone survey respondents. In our case, the size of the survey sample was determined by a financial constraint; increasing the sample size would likely produce noticeable improvements in predictive accuracy.

C. Interpreting supervised learning models

performance is reported as the average across all of the held-out folds (*42*). In our case, we repeat this entire process for all possible values of λ and α, then select the model that performs best (across held-out folds).

[3] In practice, the weighted bootstrap sample selection has little impact on results relative to a naïve selection process that evenly divides the sample into five non-overlapping sets of training (80%) and testing (20%) instances.

The original set of 5,088 features contains several behavioral metrics that are unconditionally correlated with the socioeconomic data collected in phone surveys, and a large number of features that are uncorrelated (SM Figure 3A). SM Figure 3B shows the ten features which are most highly (unconditionally) correlated with the wealth composite index; many of these features are correlated with each other, and have to do with the temporal entropy of the communications behavior of an individual's first-degree network. SM Figure 3C uses the feature partitioning described in SM Section IIC to show the distribution of the 5,088 separate R^2 values by feature type, separately for the task of predicting the composite wealth index and for the task of predicting whether the respondent owns a motorcycle. While the two sets of distributions are visually similar, the correlations are generally higher for wealth than for motorcycle ownership. Comparing the relative importance of different classes, it appears that features related to the movement patterns of an individual's social network are predictive of motorcycle ownership, whereas factors related to text messaging are most useful in predicting wealth. While it is not difficult to rationalize these observed trends ex post (for instance, it may be that text messaging is related to literacy, which is in turn correlated with wealth), we are wary of interpreting these correlations too literally.

The supervised learner described earlier optimizes the joint predictive ability of a set of features, where regularization and other methods for model selection are used to eliminate features that are not predictive or redundant. SM Figure 4A shows how model performance depends on the number of features in the model, which is in turn determined by the regularization parameter λ. SM Figure 4B illustrates how the set of features in the final model also changes as a function of the regularization parameter. When model complexity is highly penalized, few features are selected and they are initially all from the class of features that are unconditionally correlated with the response variable (in this case, the features related to the temporal entropy of the communications behavior of an individual's first-degree network.). As the penalty is reduced and more features enter the model, a more diverse set of features is selected. The optimal model includes features from a large number of different feature groups.

IV. Validation with independent sources of "ground truth" data

A. Assignment of individual mobile phone subscribers to geographic locations

Each mobile phone transaction in the call detail records is tagged with a geographic identifier corresponding to the mobile phone cell tower nearest the subscriber at the time of the transaction. Combined with a separate database containing the GPS coordinates of each cell tower, this allows us to approximately locate each individual at the time when the transaction occurs. The set of locations associated at which an individual is observed can in turn be used to infer that individuals approximate "home" location (*17*, *44*). The primary method we employ to locate an individual is to calculate the modal evening tower, defined as the single tower which the subscriber is observed to use most frequently between the hours of 8pm and 6am.[4] In developing the high-resolution visualizations (Figure 2), we additionally compute each subscriber's "center of gravity", defined as the weighted Euclidean centroid of all locations observed by the subscriber (*17*).[5] In practice, our results are not sensitive to the exact manner in which locations are computed: choosing "home" location by looking at all towers used at all hours of the day, for instance, yields nearly identical results. At the finest level of spatial granularity presented (Figure 2D), we show average locations of groups of 5-15 subscribers, where groups are determined using k-means clustering on the subscribers' centers of gravity, in order to add a layer of anonymity to the high-resolution maps.

B. Geographic aggregation: matching cell tower locations to DHS locations

When comparing the predicted wealth composite measures derived from the call records to the "ground truth" data found in the Demographic and Health Surveys, we require a comparable method of geographically aggregating data from the two sources. Our analysis uses two such levels of aggregation: district-level aggregation and "cluster"-level aggregation.

---

[4] More precisely, we compute, for each hour of the day, the most frequently used tower in that hour (the "modal tower-hour"). We then compute, for each evening, the most frequently observed modal tower-hour (the "modal tower-evening"). Finally, we compute the most frequently observed modal tower-evening across all evenings in the dataset, and use that as the subscriber's "home" location. This approach is designed to capture the location at which the subscriber spends the majority of his or her hours, rather than the location from which a majority of calls are made.

[5] Specifically, if an individual i with an modal evening tower mt_i is observed at N_i (non-unique) locations $(r_{i1}, \dots, r_{iN_i})$, we define the center of gravity as $\frac{1}{N_i}\sum_{t=1}^{N_i} r_{it}\ \ COG_i * \mathbf{1}(r_{it} - mt_i < k)$, where the indicator function restricts the weighted average to include towers within k kilometers of mt_i, to remove the influence of outliers (such as a weekend trip or short vacation). In the figures that rely on the center of gravity, we set k =10, but qualitatively similar results are obtained for a variety of reasonable thresholds (including k=∞).

When aggregating estimates at the district level, each mobile phone subscriber is assigned to a modal evening tower as described in Section IVA above. As shown in SM Figure 5, the set of unique tower locations form a voronoi division of Rwanda. We compute the average composite wealth of each voronoi division Y_v^{CDR} as the mean of the composite wealth values of all subscribers *i* whose modal evening tower is *v*, i.e. $Y_v^{CDR} = \frac{1}{N_v}\sum_{i \in v} \widehat{y_i}$, where $\widehat{y_i}$ is the predicted wealth of subscriber *i* and N_v is the number of subscribers in *v*. The average predicted composite wealth of district *d* is then computed as the weighted average of all towers falling within the district borders, $Y_d^{CDR} = \frac{1}{\sum w_{dv}}\sum_v w_{dv} * Y_v^{CDR}$, where w_{dv} indicates the proportion of the tower's voronoi cell that lies within the district boundary (SM Figure 5, inset).

Our validation estimates compare these Y_d^{CDR}, the estimates of district wealth based on mobile phone data, to the "true" wealth of the district, Y_d^{DHS}, which is computed from the DHS data as $Y_d^{DHS} = \frac{1}{\sum_{j \in d} w_j}\sum_{j \in d} w_j * y_j$, or simply the weighted average of all households *j* in district *d*, where w_j is the sampling weight given to *j* in the DHS. Y_d^{DHS} is computed separately for all households in a district, and for just the subset of household who own a mobile phone, which we later refer to as Y_d^{DHS-MP}. Correlations are weighted by population expansion factors to fit the regression line more closely to regions with large populations (*4*).

C. Cluster-level validation

We follow an analogous procedure when aggregating wealth estimates at the cluster level. Clusters are meant to approximate villages in Rwanda, and are defined in the data by the GPS locations of cluster centers collected during DHS survey collection (*31*). SM Figure 5 provides an example of how the aggregated composite wealth index is computed for a single cluster. The red dot indicates the cluster's center, and the pink shaded area represents the voronoi cell covered by the cluster. The blue dots indicate the locations of mobile phone towers, and the blue lines indicate the implied voronoi division, where dots are only shown for towers where the tower's voronoi cell overlaps with the cluster's voronoi cell. The numbers indicate w_{dv}, the proportion of the cluster's cell covered by the tower's cell. Thus, the CDR-imputed wealth value for the pink cluster will be the weighted sum of the average composite wealth predictions of each of the labelled blue cells, where the weight is given by the black number in the cell.

As noted in SM Section IC, the cluster centroids are randomly displaced by up to 10km by the DHS administrators. These displacements are intended to protect the identity of individual households, and add considerable measurement error to our ability to match DHS data to mobile phone data. The DHS documentation thus advises against disaggregating geospatial analysis below the district level (*31*).[6] For this reason, the results we emphasize in the main text that rely on DHS data use district-level aggregation.

These caveats notwithstanding, we compare phone-based estimates of average cluster wealth to DHS averages, for each of the 492 clusters in the 2010 DHS (Figure 3E). In general, the correlation at the cluster level ($r = 0.79$) is weaker than at the district level ($r = 0.92$), though for the reasons noted above this is not surprising. The primary advantage of the cluster-level analysis is that it makes it possible to analyze within-district variation, to see whether the phone-based approach picks up on differences between clusters within a district that are observed in the DHS data. SM Figure 6 thus disaggregates the results of Figure 3E by urban and rural regions. The original relationship ($r = 0.79$) is attenuated, but a correlation is still observed within both urban ($r = 0.64$) and rural ($r = 0.50$) districts.

D. Satellite night lights

Recently, a small body of work has used night-time luminosity data collected by satellites to measure economic output and growth (*45*, *46*). A key advantage of satellite data is that it is pervasive and publicly available. SI Figure 6 compares data collected by satellites on the nighttime luminosity in Rwanda with estimates of electrification based on mobile phone data. The night-light imagery, collected by the National Oceanic and Atmospheric Administration, provides a resolution of 15 arc-seconds (equivalent to a 0.74km x 0.74km grid), which is shown for the country of Rwanda (SI Figure 6A) and enlarged for the region surrounding the capital city of Kigali (SI Figure 6B). As can be seen in SI Figure 6A, there is very little variation in

[6] Excerpted from the DHS documentation (at http://dhsprogram.com/faq.cfm, accessed October 2015):
"*Can I calculate indicator estimates for areas smaller than the [district]?*
The survey design for DHS is not conducive for small area estimation. Households and respondents were selected in order to produce representative population estimates at the national and [district] level only. Any sub-[district] estimates are highly unreliable and likely to result in large standard errors.
Is it possible to do spatial analysis of DHS at the individual cluster level?
No, the sample frame is designed to ensure that the data are representative at the national and district level only."

luminosity data in poor, rural regions. Indeed, outside of the capital city of Kigali, most of the country of Rwanda appears dark and unlit.

By contrast, the approach based on phone data captures a great deal of variation even in the most rural parts of the country, and allows for the distinction between households that have access to electricity and households that are brightly lit at night (Figure 3). We use the method described in the paper to predict how each of the 1.5 million subscribers would respond to the survey question, "Does your household have electricity?" using methods analogous to those used to predict composite wealth. Each subscriber's center of gravity is used to place the individual in a grid cell, and the average predicted response is computed across all subscribers. These values are then used to construct a map of predicted electrification in the Kigali region (SI Figure 6C). While the two images are visually similar, they are designed to capture slightly different phenomena: the night light imagery is optimized "to observe dim signals such as city lights, gas flares, auroras, wildfires, and reflected moonlight"; the mobile phone-based predictions are constructed to map household electrification. In urban settings like Kigali, we presume these to be strongly correlated, but in more rural regions the distinction is more important.

V. Generalizability and external validity

The results in Figure 1 illustrate how our method can be used to infer individual characteristics (in our case, phone survey responses) from passively-generated transactional data (mobile phone records), for the population of individuals who generate such data (the population of active mobile phone subscribers). This method, we believe, should generalize to a wide range of contexts where it is possible to supplement large transactional datasets with targeted surveys. SM Section VI provides several examples of possible applications of this method that extend far beyond the population of Rwandan mobile phone owners, which we hope we and other researchers can improve upon in future work.

A. Population inference from a sample of mobile phone subscribers

The model fit on the sample of 856 respondents is then used to generate out-of-sample predictions for the population of 1.5 million mobile phone subscribers in Rwanda. To validate the accuracy of these predictions, we compare the aggregated output of this model to DHS data

aggregated at the same geographic level. In performing this validation, we observe two distinct results.

First, we find that the average wealth of a district, as predicted by the mobile phone data (Y_d^{CDR}), is strongly correlated (r = 0.917) with the average wealth of mobile-phone owning households in that district (Y_d^{DHS-MP}), as reported in the 2010 DHS.[7] This provides objective validation that our method can reconstruct the distribution of wealth of a population for whom we expect it to be representative, i.e., mobile phone owners. Since our estimate of Y_d^{CDR} was constructed "in a vacuum" and without access to the DHS data, there is no possibility that the relationship is mechanical or that the model was overfit to the DHS target.

Second, as shown in Figure 3, we observe an equally strong correlation (r = 0.916) between the phone-based estimates of district wealth (Y_d^{CDR}) and the average wealth of all households in the district (Y_d^{DHS}). This result indicates that, at least in Rwanda, our method can approximate the distribution of wealth of the full national population. This is true despite the fact that Y_d^{DHS} is constructed from a sample that is representative of the population of all Rwandans, while Y_d^{CDR} is constructed from a sample that is representative of the population of active mobile phone subscribers. And it is true despite the fact that, as we have shown in prior work (*30*), these two populations are different: mobile phone subscribers in general are wealthier, better educated, and more likely to be male.

B. Generalizing to other contexts

In other contexts, it is possible that one could accurately reconstruct the wealth of phone owners from phone records (as we do in Figure 1), but not be able to accurately reconstruct the distribution of wealth of the full population from out-of-sample inferences about mobile subscribers (as we do in Figure 3). In the general case, assume the researcher has conducted a targeted survey with a sample of individuals (POP^{survey}), who we assume are a random, representative sample of the population of individuals for whom the researcher has transactional data (POP^{data}),[8] who in turn constitute a subset of the full population (POP^{full}). As a broad

[7] In this DHS, mobile phones are owned by approximately 42% of the sample or 5,315 households.
[8] Our efforts to ensure to draw a sample for POP^{survey} that was representative of POP^{data} are described in SM Section 1A.

heuristic, the more representative POP^{data} are of POP^{full}, the more effective we expect this approach will be; if POP^{data} are not representative, then validating estimates against external data on POP^{full}, as we have with Figure 3, is a critical step.

In Rwanda, there are several possible explanations for why we are able to reconstruct the distribution of wealth of POP^{full} from POP^{data} even though we know the latter is not a representative sample of the former. The simplest explanation, however, is the fact that in Rwanda, Y_d^{DHS-MP} is closely correlated with Y_d^{DHS} (r = 0.984). In other words, there exists a strong correlation between the average wealth of region's population and the average wealth of a region's mobile phone-owning population. In situations where the selection process into mobile phone ownership is uniform across regions, this result is likely to generalize.

More broadly, as mobile phones are quickly adopted in developing countries (*11*), it may become more tenable to predict wealth and poverty from mobile phone data in a broad range of geographic contexts. In general, however, POP^{data} may not be representative of POP^{full}, and the ability to infer properties of POP^{full} from POP^{data} will depend heavily on the context of the application. In Rwanda, for instance, our analysis was facilitated by the unusual concentration of the mobile phone market. In more fragmented markets, the approach might need to be adapted if there is systematic selection of subscribers into mobile phone network providers, unless the researcher can obtain data from all relevant operators.[9] Similarly, the near-ubiquitous coverage and high density of cellular towers in Rwanda (SM Figure 5) made it possible to include remote regions in POP^{survey}, which in turn allowed us to construct high-resolution estimates for the entire country (Figure 2).

Related, our analysis focuses on predicting the composite wealth of a subscriber ($\widehat{y_i}$), where the composite wealth is defined the first principal component of the assets and characteristics of the household. This target variable was well-suited to the Rwandan context, where many phones are shared within households (*30*), income is typically pooled among household members, and the majority of households rely on subsistence agriculture. In other contexts, where phone use is more individual and it is more common to earn a fixed wage, individual income may be a more

[9] Here, an intriguing possibility is governments would require, or other institutions would provide incentives, to operators to make data available for humanitarian use (*47*).

natural target prediction variable. However, one limitation of the approach we propose is that it is designed to model response variables that can be elicited through short, structured phone interviews. Thus, it would be difficult to use this method to predict consumption or expenditures, which typically require extensive survey modules, or more sensitive topics that respondents do not feel comfortable discussing over the phone.

Other idiosyncrasies of the Rwandan context, such as the dominance of pre-paid accounts and the per-second billing structure, likely impacted the set of features engineered and selected through supervised learning. A fragmented market would also affect the model fit on POP^{survey}, as a single operator's call detail records would only capture partial information for a competitor's subscribers. However, we do not expect that such idiosyncrasies would necessarily weaken one's ability to train a model on POP^{survey}, or imply non-representativeness of POP^{data}. In other words, while the fitted model would change, the process for fitting the model would remain the same, and any changes in goodness of fit are hard to predict ex ante.

VI. Applications and extensions

The focus of this paper has been on predicting poverty and wealth from mobile phone data. However, with minimal changes, an analogous approach could be used to predict a much broader set of characteristics (not just wealth and poverty) by supplementing other large datasets (not just mobile phone records) with other targeted data collection (not just phone surveys). We conclude with a discussion of several ways in which the methods presented in this paper could be further extended.

A. Interim national statistics

One compelling use case for the phone-based predictions of poverty and wealth is as a source of interim national statistics. The thought experiment we imagine is a policymaker who needs to make a decision that requires knowledge of the distribution of wealth. If the policymaker does not have the resources to collect original data, in many countries she would likely rely on data from the most recent nationally-representative survey. As we have noted in the main text, in many developing countries, such data is woefully out of date (*3*).

Rwanda, in this sense, is unrepresentative of much of sub-Saharan Africa, as multiple nationally-representative surveys have been conducted in Rwanda in recent years. Even so, if our policymaker were in Rwanda in 2010, it is likely that she would use the results of the 2007 DHS, as the results from the 2010 DHS were not made public until mid-2011. As can be seen in SM Figure 8, however, the correlation between estimates of wealth based on mobile phone data and 2010 DHS data ($r = 0.91$) is in fact greater than the correlation between the two successive rounds of DHS data ($r = 0.84$). Thus, if she were to use the 2007 DHS data to identify the districts with below-average wealth, as defined by the first principal component of 2007 DHS assets, she would correctly identify 14 of the 20 districts (70%) which had below-average wealth in 2010, defined by the first principal component of 2010 DHS assets. By contrast, if she were to use the estimates of district wealth compute from the call records, she would correctly identify 17 of the 20 districts (85%). In countries where longer lags exist between successive survey waves, these differences could be quite meaningful.

B. Targeting individuals for welfare subsidies, critical information, or promotions

The method we describe makes it possible to predict the characteristics of millions of individual mobile phone subscribers. This creates obvious opportunities for profit, if firms wish to target advertising or promotional content to specific demographics. It may also facilitate new methods for targeting target resources to individuals with the greatest need, or for providing information to individuals likely to be at risk. As currently developed, the method focuses on predicting a composite asset index, but in principle a similar approach could be used to estimate consumption as in a proxy means test (*48*). Relative to the more common asset-based proxy means test, a method based on phone (or other digital transactions) data has certain advantages: it could be targeted to individuals rather than to households; the observed characteristics, derived from call data, can be observed with little marginal cost once the fixed cost of data access is paid; the highly nonparametric process for fitting the target variable to observed metrics could allow for more accurate targeting; and the allocation rule could be made difficult to game.

Yet any implementation of such a system will also face significant obstacles. Many individuals, and particularly the most vulnerable, still do not generate a digital transaction log, and may remain “off the grid” for the foreseeable future. Even if the goal were to only reach mobile phone owners, there would be significant barriers to obtaining the necessary data on phone use,

particularly in markets with multiple operators. Finally, as we discuss below, it is likely that the function mapping phone use to the target variable will change over time, either through natural shifts in patterns of device use or through deliberate actions of individuals who wish to alter their behavior to become eligible for benefits. One can imagine possible solutions to these challenges – for example by distributing phones to potential beneficiaries, government-mandated data sharing regulations, or frequent model rebasing – but the path forward is not trivial.

C. Measuring changes over time, and impact evaluation

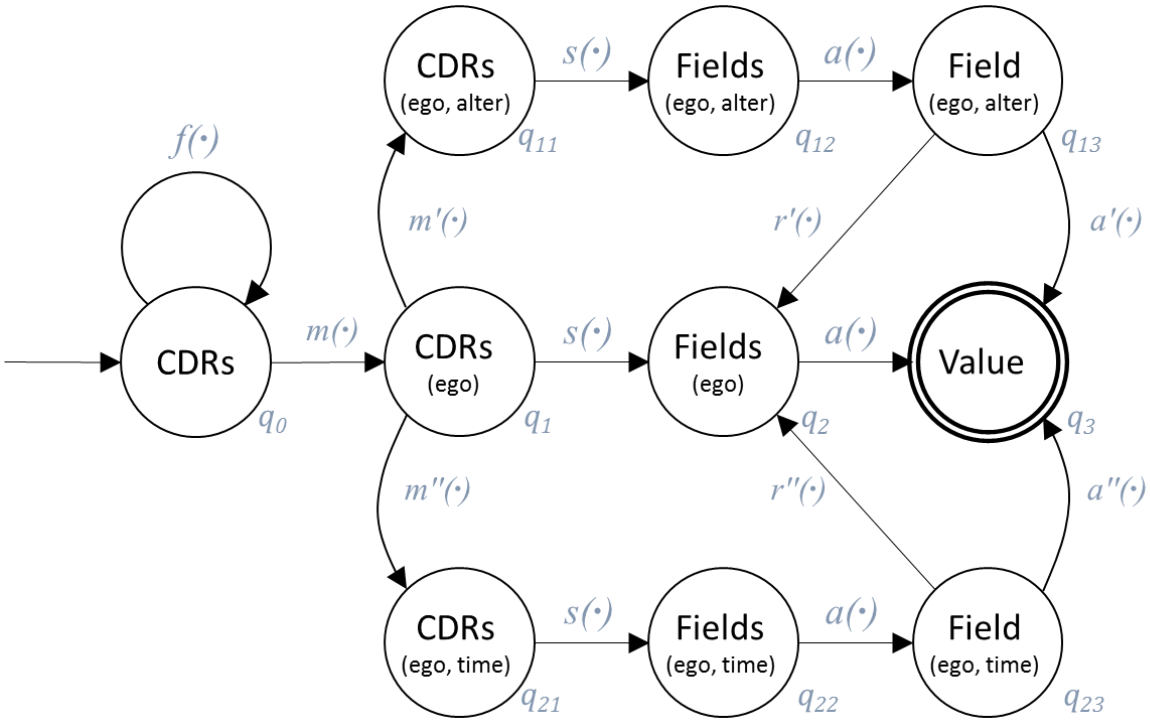
Perhaps most compelling, and also most speculative, is the possibility that related methods could be used to detect changes over time in the social, economic, or mental state of an individual or small region. A large body of work indicates that events in the real world have unique fingerprints in transactional data (*6*, *49*, *50*), and it is easy to imagine that a sudden period of hunger or a bout of depression would be manifest in the phone records of the affected. If a derivative approach could be used to reliably estimate changes in welfare over time, it would enable new approaches to impact evaluation and program monitoring, among other applications.

As we have stated repeatedly, however, we do not assume that a model trained on a specific population at a specific point in time could be used to draw inferences about a different population or a different time period. Rather, we expect that the true mapping from digital data to welfare outcomes is context-dependent, and that the model estimated in one time period would deteriorate as time passes from the moment at which it is fit (*51*). An interesting avenue to pursue here would be to periodically rebase the model by conducting additional surveys to update the model parameters, possibly using online machine learning methods to determine when new surveys are needed and with which populations.

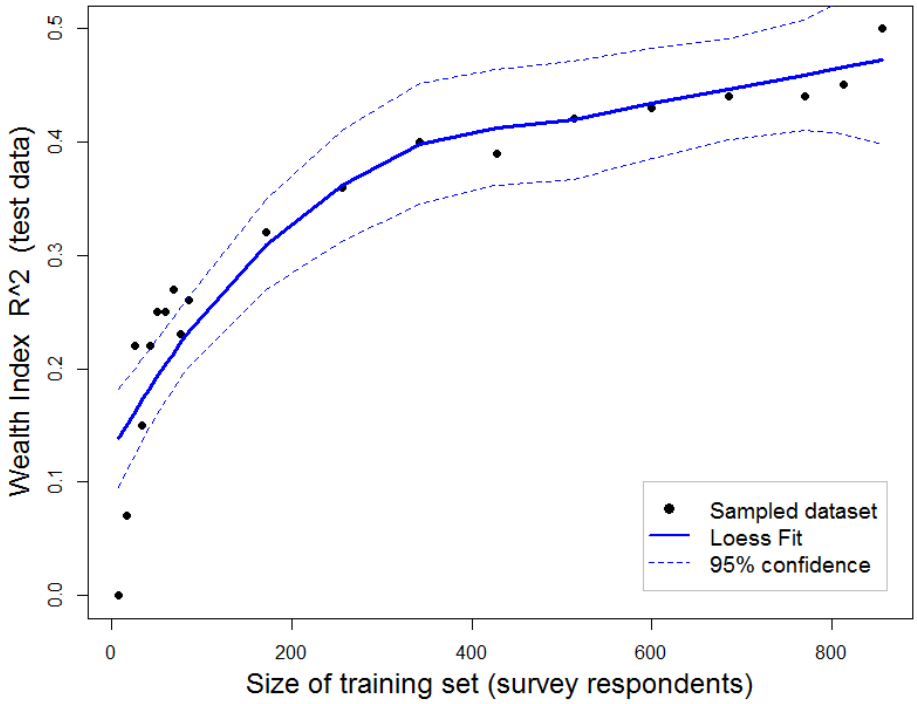
A	Elastic Net		Random Forest	
	r	R^2	r	R^2
Optimal DFA-based model	0.68	0.46	0.63	0.40
"Intuitive" 5-feature model	0.44	0.20	0.37	0.14
Single-feature model	0.61	0.38	0.46	0.22

B	Accuracy	AUC	F score	Baseline
Owns a refrigerator	0.75	0.88	0.40	0.11
Household has electricity	0.72	0.85	0.74	0.60
Owns a television	0.73	0.84	0.72	0.49
Owns a bicycle	0.64	0.68	0.47	0.30
Owns a motorcycle/scooter	0.72	0.67	0.20	0.11
Owns a radio	0.92	0.50	0.96	0.96

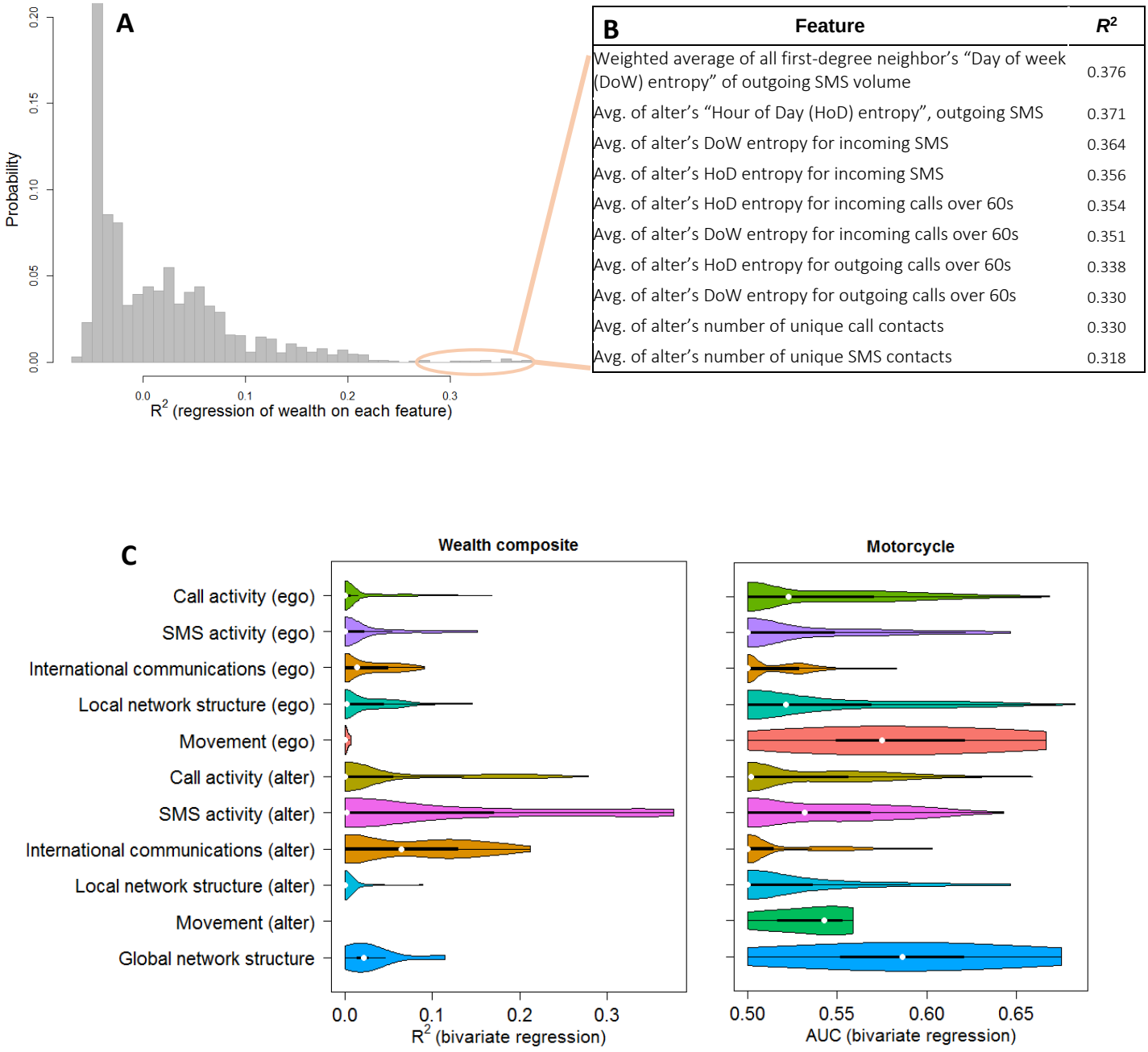
Table S1. Cross-validated performance of predictive models. The models are trained using 5-fold cross-validation on the set of 856 survey respondents. **(A)** Measures of goodness of fit (correlation coefficient and R^2) for two optimized models: the elastic net which selects 101 features, and a random forest regressor. For comparison, we show performance measures trained on set of five features commonly cited in the literature (total call volume, total SMS volume, total international call volume, radius of gyration, degree centrality); and for a model with the single most predictive feature (the weighted average of all first-degree neighbor's "Day of week (DoW) entropy" of outgoing SMS volume). **(B)** Performance measures and a naïve baseline for predicting binary survey responses. Accuracy indicates the fraction of correct predictions from regularized logistic regression; Area under curve (AUC) indicates the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one, which helps account for the fact that some assets are quite common while others are quite uncommon; the F score provides a performance measure that balances precision and recall; the Baseline is the fraction of respondents who report owning the asset.



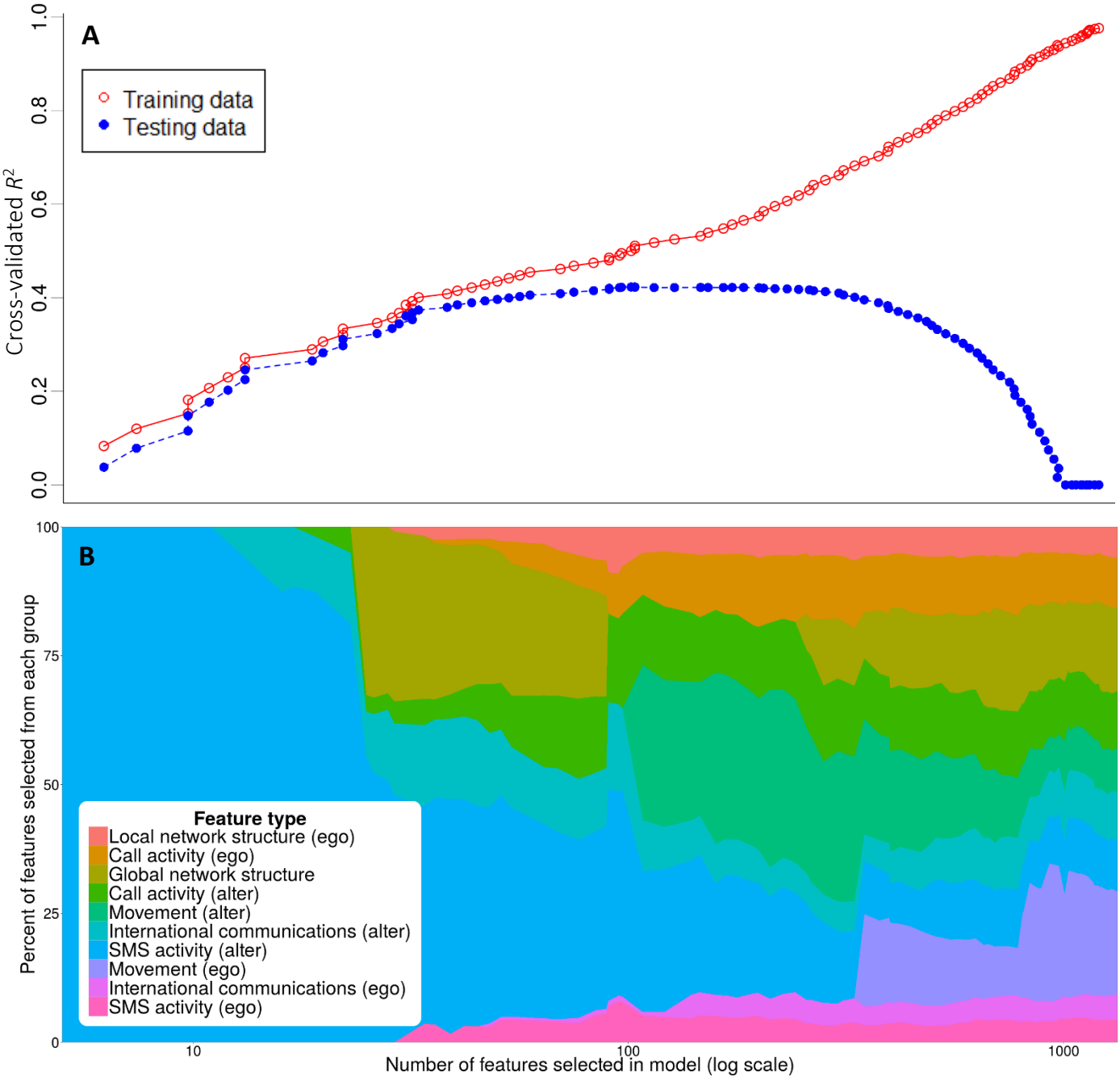
SM Fig. 1. Deterministic finite automaton used for feature engineering. Circles represent states and arrows represent legal transitions, where q_0 is the start state and q_3 is the accepting (end) state. The final output from the DFA is a single numerical value, which is equivalent to a single behavioral metric, or "feature."



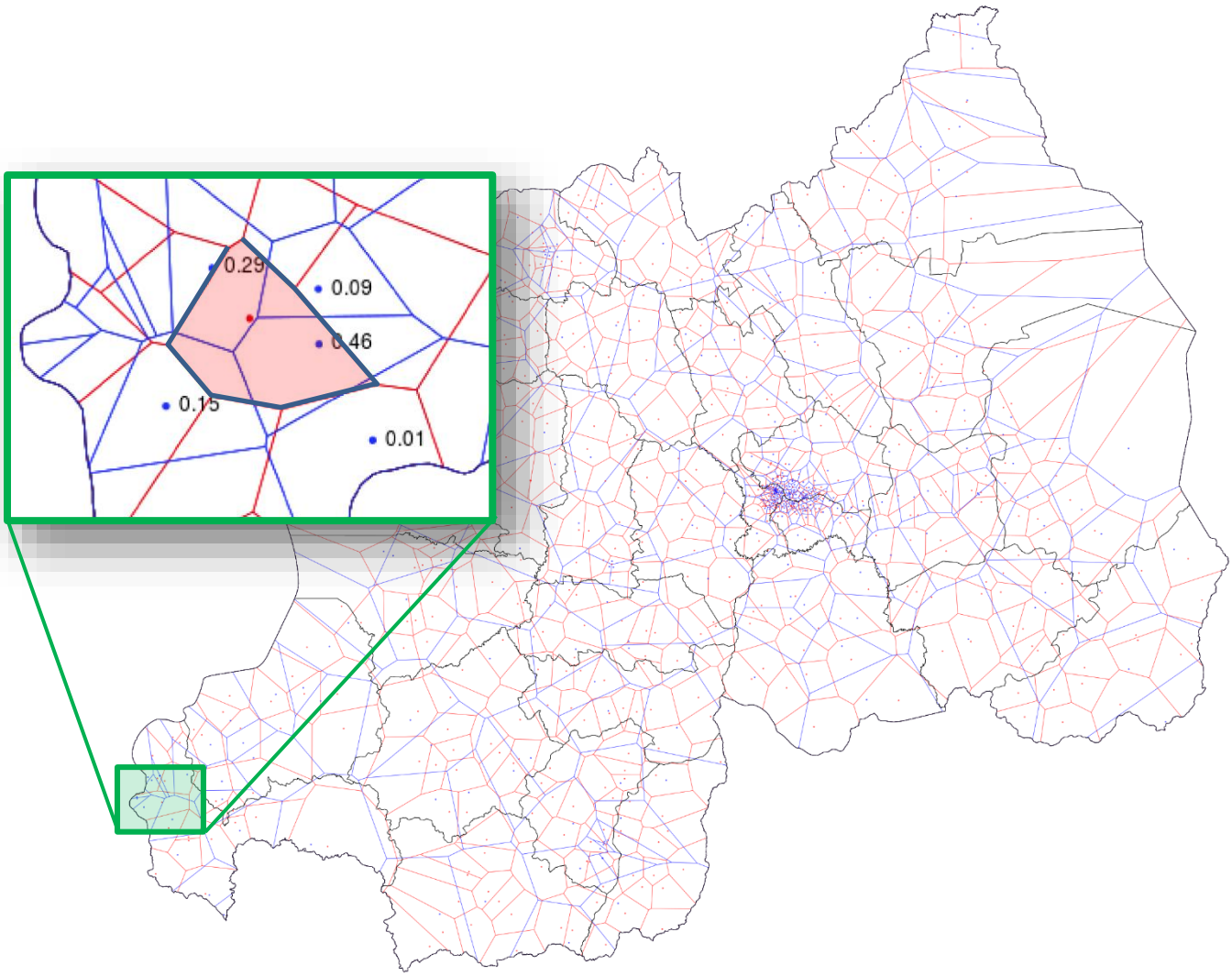
SM Fig. 2. Model performance. As the number of training instances increases, the performance of the model steadily improves. Adding additional respondents would likely enable continued increases in predictive accuracy.



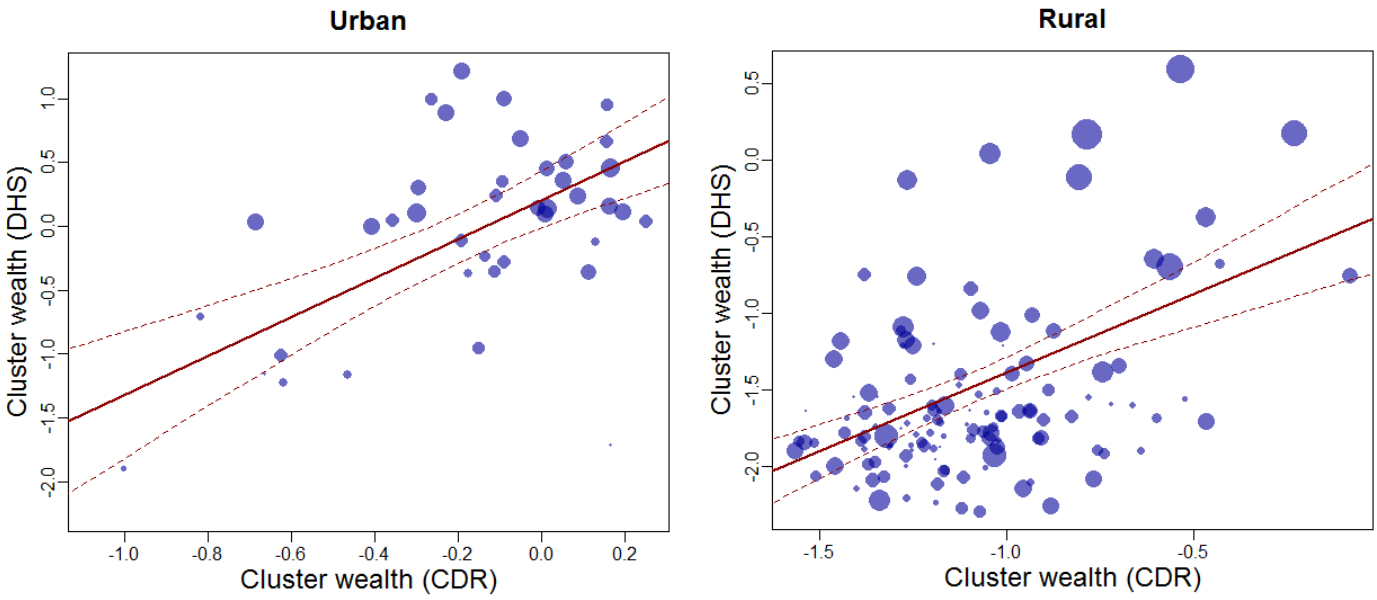
SM Fig. 3. Metrics of phone use that correlate with survey responses. (A) The distribution of R^2 values from 5,088 separate regressions of the wealth composite index on each feature, showing average accuracy on the test set after 5-fold cross validation. **(B)** Representative list of features strongly correlated with the composite wealth index. **(C)** Distribution of R^2 values by feature class, for different response variables.



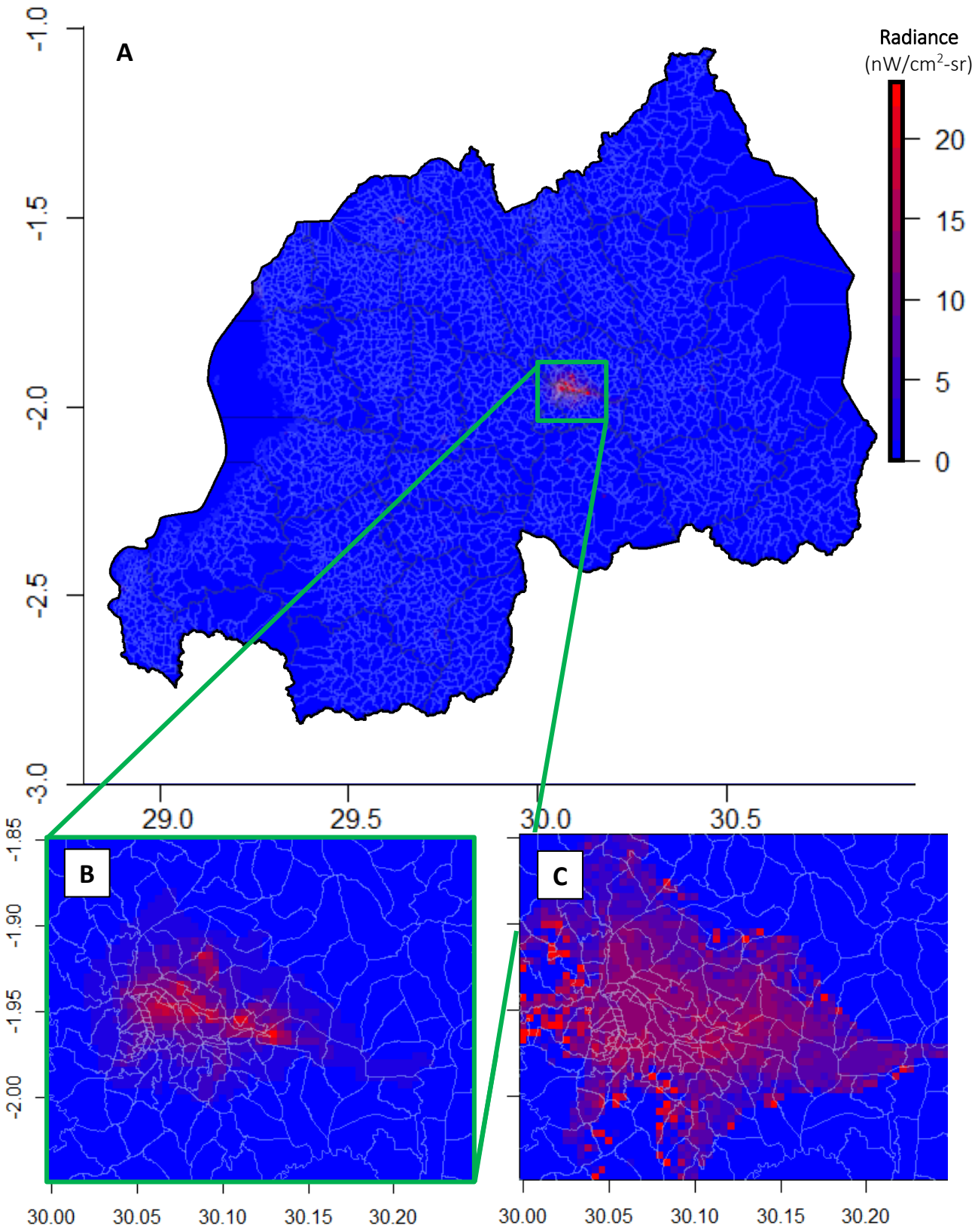
SM Fig. 4. The impact of regularization on model performance and feature selection. (A) Average cross-validated performance, showing average R^2 across 5 random training folds and testing folds. Increasing the regularization parameter produces more parsimonious models with fewer features. The optimal regularized model includes 101 features. Including additional features causes the model to overfit on the set of training instances, while excluding features degrades predictive accuracy. **(B)** Composition of features selected for models of varying complexity. When model complexity is highly penalized, few features are selected and they are all initially from the same class (SMS activity of the ego's first-degree network of "alters"). As the penalty is reduced and more features enter the model, a more diverse set of features is selected.



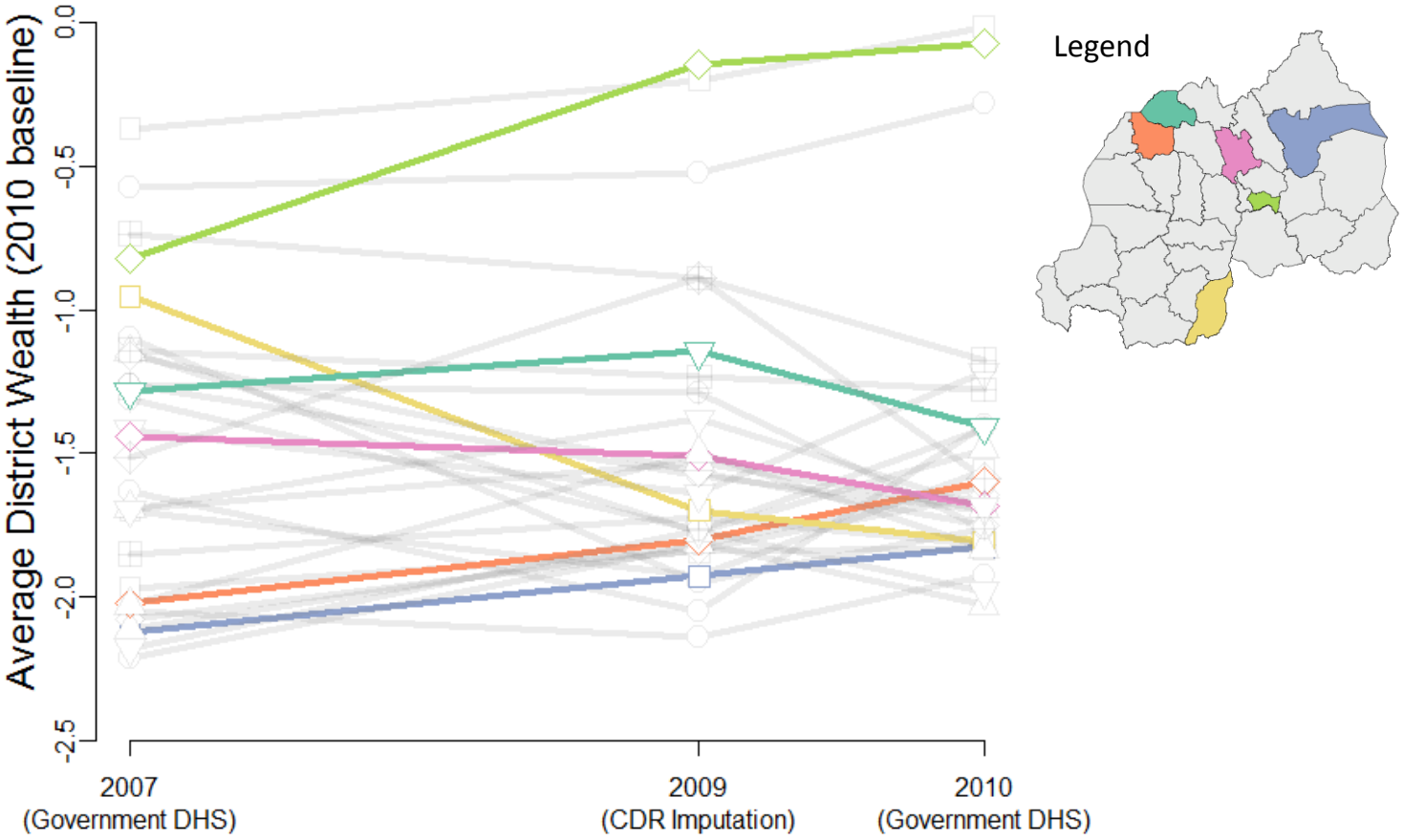
SM Fig. 5. Matching locations of mobile phone subscribers to geographic regions in household survey data. Rwanda is comprised of 30 administrative districts, shown with black borders. In 2009, Rwanda contained roughly 300 unique mobile phone towers, indicated with blue dots. The voronoi tessellation formed by these towers is shown with blue lines. The 2010 DHS sample frame used 492 clusters, the centroids of which are indicated with red dots, and the voronoi tessellation with red lines. The inset figure illustrates how the areas of overlap between the two voronoi divisions are used to compare information aggregated within mobile phone towers to information aggregated within DHS clusters.



SM Fig. 6. Comparison of wealth predictions to government survey data, separately for urban and rural areas. The left figure restricts the analysis to DHS clusters within the urban capital of Kigali; the right panel includes only clusters outside of Kigali. Solid and dashed red lines indicate the regression line and 95% confidence intervals.



SM Fig. 7. Comparison of satellite night-light data to phone-based estimates of electrification. (A) Map of Rwanda showing night-time luminosity, as captured by satellites orbiting the earth (NOAA National Geophysical Data Center). **(B)** Enlargement of satellite imagery in the region near Kigali, the capital of Rwanda. **(C)** Predicted household electrification, based on call records, using a model fit on how 856 survey respondents answered the question, "Does your household have electricity?" and projected onto the full population of mobile subscribers.



SM Fig. 8. Phone-based wealth predictions accurately interpolate between traditional rounds of household surveys. Each of Rwanda's 30 districts is represented as a line, where the values in 2007 and 2010 are calculated using household survey data from the Rwandan Demographic and Health Surveys (DHS) of 7,377 and 12,792 households, respectively. The value in 2009 is computed from the mobile phone call detail records (CDR) of roughly 1.5 million subscribers in Rwanda, using a predictive model calibrated on a sample of 856 survey respondents. Every fifth district (ordered by predicted wealth in 2009) is colored to highlight changes over time of six different districts.

References and Notes

1. S. Kuznets, Economic growth and income inequality. *Am. Econ. Rev.* **45**, 1–28 (1955).

2. G. S. Fields, Changes in poverty and inequality in developing countries. *World Bank Res. Obs.* **4**, 167–185 (1989). doi:10.1093/wbro/4.2.167

3. M. Jerven, *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It* (Cornell Univ. Press, Ithaca, NY, 2013).

4. C. Elbers, J. O. Lanjouw, P. Lanjouw, Micro-level estimation of poverty and inequality. *Econometrica* **71**, 355–364 (2003). doi:10.1111/1468-0262.00399

5. M. Ghosh, J. N. K. Rao, Small area estimation: An appraisal. *Stat. Sci.* **9**, 55–76 (1994). doi:10.1214/ss/1177010647

6. D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M. Van Alstyne, Computational social science. *Science* **323**, 721–723 (2009). Medline doi:10.1126/science.1167742

7. G. King, Ensuring the data-rich future of the social sciences. *Science* **331**, 719–721 (2011). Medline doi:10.1126/science.1197872

8. N. Eagle, M. Macy, R. Claxton, Network diversity and economic development. *Science* **328**, 1029–1031 (2010). Medline doi:10.1126/science.1186605

9. H. Choi, H. Varian, Predicting the present with Google Trends. *Econ. Rec.* **88**, 2–9 (2012). doi:10.1111/j.1475-4932.2012.00809.x

10. W. Wang, D. Rothschild, S. Goel, A. Gelman, Forecasting elections with non-representative polls. *Int. J. Forecast.* **31**, 980–991 (2015). doi:10.1016/j.ijforecast.2014.06.001

11. "The mobile economy 2014" (GSMA Intelligence, 2014); www.gsmamobileeconomy.com/GSMA_ME_Report_2014_R2_WEB.pdf.

12. J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, A.-L. Barabási, Uncovering individual and collective human dynamics from mobile phone records. *J. Phys. A* **41**, 224015 (2008). doi:10.1088/1751-8113/41/22/224015

13. J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A. L. Barabási, Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332–7336 (2007). Medline doi:10.1073/pnas.0610245104

14. G. Palla, A. L. Barabási, T. Vicsek, Quantifying social group evolution. *Nature* **446**, 664–667 (2007). Medline doi:10.1038/nature05670

15. M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008). Medline doi:10.1038/nature06958

16. X. Lu, E. Wetter, N. Bharti, A. J. Tatem, L. Bengtsson, Approaching the limit of predictability in human mobility. *Sci. Rep.* **3**, 2923 (2013). Medline doi:10.1038/srep02923

17. J. E. Blumenstock, Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Inf. Technol. Dev.* **18**, 107–125 (2012). doi:10.1080/02681102.2011.643209

18. V. Frias-Martinez, J. Virseda, in *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development* (Association for Computing Machinery, New York, 2012), pp. 76–84; http://doi.acm.org/10.1145/2160673.2160684.

19. P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, A. J. Tatem, Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15888–15893 (2014). Medline doi:10.1073/pnas.1408439111

20. G. C. Cawley, N. L. C. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **11**, 2079–2107 (2010).

21. D. Filmer, L. H. Pritchett, Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography* **38**, 115–132 (2001). Medline

22. J. Blumenstock, N. Eagle, Divided we call: Disparities in access and use of mobile phones in Rwanda. *Inf. Technol. Int. Dev.* **8**, 1–16 (2012).

23. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **67**, 301–320 (2005). doi:10.1111/j.1467-9868.2005.00503.x

24. L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, *Classification and Regression Trees* (Chapman and Hall/CRC Press, New York, ed. 1, 1984).

25. B. Abelson, K. R. Varshney, J. Sun, in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, 2014), pp. 1563–1572; http://doi.acm.org/10.1145/2623330.2623335.

26. National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], ICF International, "Rwanda Demographic and Health Survey 2010," *DHS Final Reports* (publication ID FR259, NISR, MOH, and ICF International, Calverton, MD, 2012).

27. M. Jerven, "Benefits and costs of the data for development targets for the post-2015 development agenda," in *Data for Development Assessment Paper* (Copenhagen Consensus Center, 2014).

28. Y.-A. de Montjoye, L. Radaelli, V. K. Singh, A. S. Pentland, Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* **347**, 536–539 (2015). Medline doi:10.1126/science.1256297

29. A. Wesolowski, C. O. Buckee, L. Bengtsson, E. Wetter, X. Lu, A. J. Tatem, Commentary: Containing the ebola outbreak - the potential and challenge of mobile network data. *PLOS Curr.* 10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e (2014). Medline doi:10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e

30. J. Blumenstock, N. Eagle, "Mobile divides: Gender, socioeconomic status, and mobile phone use in Rwanda," in *Proceedings of the 4th ACM/IEEE International Conference on*

Information and Communication Technologies and Development (Association for Computing Machinery, New York, 2010), article no. 6; http://doi.acm.org/10.1145/2369220.2369225.

31. Ministry of Health (MOH) [Rwanda], National Institute of Statistics of Rwanda (NISR), and ICF Macro, "Rwanda DHS, 2007-08 - Rwanda Interim Demographic and Health Survey (English)," *DHS Final Reports* (publication ID FR215, MOH, NISR, and ICF Macro, Calverton, MD, 2009).

32. A. Deaton, S. Zaidi, *Guidelines for Constructing Consumption Aggregates for Welfare Analysis* (World Bank Publications, 2002).

33. H.-P. Kriegel, P. Kröger, E. Schubert, A. Zimek, "A general framework for increasing the robustness of PCA-based correlation clustering algorithms," in *Scientific and Statistical Database Management*, B. Ludäscher, N. Mamoulis, Eds. (Lecture Notes in Computer Science Series, Springer, Berlin, Heidelberg, 2008), pp. 418–435.

34. D. Hillger, T. Kopp, T. Lee, D. Lindsey, C. Seaman, S. Miller, J. Solbrig, S. Kidder, S. Bachmeier, T. Jasmin, T. Rink, First-light imagery from Suomi NPP VIIRS. *Bull. Am. Meteorol. Soc.* **94**, 1019–1029 (2013). doi:10.1175/BAMS-D-12-00097.1

35. V. Frias-Martinez, J. Virseda, E. Frias-Martinez, "Socio-economic levels and human mobility," paper presented at the QualMeetsQuant Workshop at the 4th International Conference on Information and Communication Technologies and Development, London, 13 to 16 December 2010.

36. J. Blumenstock, Y. Shen, N. Eagle, "A method for estimating the relationship between phone use and wealth," paper presented at the QualMeetsQuant Workshop at the 4th International Conference on Information and Communication Technologies and Development, 13 to 16 December 2010.

37. A. Decuyper *et al*., http://arxiv.org/abs/1412.2595 (2014).

38. T. Gutierrez, G. Krings, V. D. Blondel, http://arxiv.org/abs/1309.4496 (2013).

39. M. O. Rabin, D. Scott, Finite automata and their decision problems. *IBM J. Res. Develop.* **3**, 114–125 (1959). doi:10.1147/rd.32.0114

40. V. D. Blondel, A. Decuyper, G. Krings, http://arxiv.org/abs/1502.03406 (2015).

41. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).

42. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics, Springer, New York, 2009), vol. 2.

43. B. Efron, R. Tibshirani, Improvements on cross-validation: The 632+ bootstrap method. *J. Am. Stat. Assoc.* **92**, 548–560 (1997).

44. R. Ahas, S. Silm, O. Järv, E. Saluveer, M. Tiru, Using mobile positioning data to model locations meaningful to users of mobile phones. *J. Urban Technol.* **17**, 3–27 (2010). doi:10.1080/10630731003597306

45. J. V. Henderson, A. Storeygard, D. N. Weil, Measuring economic growth from outer space. *Am. Econ. Rev.* **102**, 994–1028 (2012). Medline doi:10.1257/aer.102.2.994

46. X. Chen, W. D. Nordhaus, Using luminosity data as a proxy for economic statistics. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8589–8594 (2011). Medline doi:10.1073/pnas.1017031108

47. Y.-A. de Montjoye, J. Kendall, C. F. Kerry, "Enabling humanitarian use of mobile phone data," in *Issues in Technology Innovation* (Brookings Center for Technology Innovation, 2014); http://dspace.mit.edu/handle/1721.1/92821.

48. V. Alatas, A. Banerjee, R. Hanna, B. A. Olken, J. Tobias, Targeting the poor: Evidence from a field experiment in Indonesia. *Am. Econ. Rev.* **102**, 1206–1240 (2012). Medline doi:10.1257/aer.102.4.1206

49. J. P. Bagrow, D. Wang, A.-L. Barabási, Collective response of human populations to large-scale emergencies. *PLOS ONE* **6**, e17680 (2011). Medline doi:10.1371/journal.pone.0017680

50. J. E. Blumenstock, N. Eagle, M. Fafchamps, "Risk sharing and mobile phones: Evidence in the aftermath of natural disasters," Working paper (2014); www.jblumenstock.com/files/papers/jblumenstock_mobilequakes.pdf.

51. D. Lazer, R. Kennedy, G. King, A. Vespignani, The parable of Google Flu: Traps in big data analysis. *Science* **343**, 1203–1205 (2014). Medline doi:10.1126/science.1248506