Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being[1]

Ryan Engstrom[2]

Jonathan Hersh[3]

David Newhouse[4]

This draft: April, 2017

Abstract

Estimating local poverty is important for targeting aid and forming policy in developing economies, but is severely constrained by the expense of collecting household survey and census data. This paper investigates the ability of features derived from high spatial resolution satellite images to accurately estimate poverty and economic well-being. Both object and texture features are extracted from satellite images of Sri Lanka, which are used to estimate poverty rates and average log consumption for 1,291 villages (Grama Niladhari Divisions). Features include the number and density of buildings, the prevalence of shadow area, which is a proxy for building height, the number of cars, density and length of roads, type of farmland, and roof material. A simple linear model using only these features as explanatory variables, developed using Lasso regularization to prevent overfitting, explains nearly sixty percent of both poverty headcount rates and average log consumption at the village level. Predictions remain accurate throughout the village welfare distribution. Buildings, shadows, and road characteristics have strong associations with poverty. Two sample applications, extrapolating predictions into adjacent areas and estimating local area poverty using an artificially reduced census, confirm the out of sample predictive capabilities. We conclude that high spatial resolution imagery has the potential to transform small area poverty estimation, with implications for optimal survey design, and that the use of derived features offers important advantages to existing methods that employ satellite imagery to estimate poverty.

Keywords: poverty estimation, satellite imagery, machine learning
JEL classification: I32, C50

---

[1] This project benefited greatly from discussions with Sarah Antos, Ana Areias, Marianne Baxter, Sam Bazzi, Azer Bestavros, Kristen Butcher, John Byers, Francisco Ferreira, Ray Fisman, Alex Guzey, Klaus-Peter Hellwig, Kristen Himelein, Tariq Khokhar, Trevor Monroe, Dilip Mookherjee, Pierre Perron, Bruno Sánchez-Andrade Nuño, Kiwako Sakamoto, David Shor, Benjamin Stewart, Andrew Whitby, Nat Wilcox, Nobuo Yoshida and seminar participants at the Boston University Development Reading Group, Chapman University, The World Bank, and the Department of Census and Statistics of Sri Lanka. All remaining errors in this paper remain the sole responsibility of the authors. Sarah Antos, Benjamin Stewart, and Andrew Copenhaver provided assistance with texture feature classification. Object imagery classification was assisted by James Crawford, Jeff Stein, and Nitin Panjwani at Orbital Insight, and Nick Hubing, Jacqlyn Ducharme, and Chris Lowe at Land Info, who also oversaw imagery pre-processing. Hafiz Zainudeen helped validate roof classifications in Colombo. Colleen Ditmars and her team at DigitalGlobe facilitated imagery acquisition, Dung Doan and Dilhanie Deepawansa developed and shared the census-based poverty estimates, and we thank Dr. Amare Satharasinghe for authorizing the use of the Sri Lankan census data. Liang Xu provided research assistance. Zubair Bhatti, Benu Bidani, Christina Malmberg-Calvo, Adarsh Desai, Nelly Obias, Dhusynanth Raju, Martin Rama, and Ana Revenga provided additional support and encouragement. The authors gratefully acknowledge financial support from the Strategic Research Program and World Bank Big Data for Innovation Challenge Grant, and the Hariri Institute at Boston University. The views expressed here do not necessarily reflect the views of the World Bank Group or its executive board, and should not be interpreted as such.
[2] rengstro@gwu.edu Department of Geography, George Washington University, 1922 F Street, Washington DC
[3] hersh@chapman.edu Argyros School of Business, Chapman University, 1 University Dr., Orange, CA
[4] dnewhouse@worldbank.org, Poverty and Equity Global Practice, World Bank, 1818 H Street, Washington DC

1 Introduction

Despite the best efforts of national statistics offices and the international development community, small area estimates of poverty and economic welfare remain rare. This is due in part to the lack of available household survey data measuring economic welfare in developing countries. Between 2002 and 2011, as many as 57 countries conducted zero or only one survey capable of producing poverty statistics, and data are scarcest in the poorest countries. (Serajuddin et al, 2015). But even in countries where data are collected regularly, household surveys are typically too small to produce reliable estimates below the district level. Generating welfare estimates for smaller areas requires both a household welfare survey and contemporaneous census data, and the latter is typically available once per decade at best. Furthermore, safety concerns prohibit survey data collection in many conflict areas altogether. Lack of timely information on living standards in small areas impedes the efforts of policymakers and aid organizations to direct scarce resources to the poor, and prevents their constituents to hold them accountable for doing so.

Satellite imagery has generated considerable enthusiasm as a potential supplement to household data that can help fill these severe data gaps. In recent years, private companies such as DigitalGlobe and Airbus have rapidly expanded the coverage and availability of high spatial resolution imagery (HSRI), driving down commercial prices. Planet (formerly Planetlabs) currently operates more satellites than any organization other than the US and Russian governments, and just recently, successfully launched 88 dove satellites that will allow for coverage of the entire globe with imagery resolution of 4 to 5 m per pixel on a daily basis. Continued technological advances will increasingly allow social scientists to benefit from this type of imagery, which has been utilized intensively by the intelligence and military communities for decades.

This paper investigates the ability of object and texture features derived from HSRI (High Spatial Resolution Imagery) to estimate and predict poverty rates at local levels. The area of our study covers 3,500 square kilometers in Sri Lanka, which contain 1,291 villages (Grama Niladhari (GN) divisions). For each village, we extract both object and "texture" feature to use as explanatory variables in poverty prediction models. Object features extracted include the number of cars, number and size of buildings, type of farmland (plantation vs. paddy), the type of roofs, the share of shadow pixels (building height proxy), road extent and road material, along with textural measures. These features are identified using a combination of deep learning based Convolutional Neural Networks (CNN) and eCognition, an object based image processing approach. The eCognition objects were edited and in some cases created using manual "heads up" digitization. Additionally, texture features, values derived using computer vision approaches that characterize the spatial variability in an area or neighborhood within an image, were also calculated. These satellite derived features were then matched to household estimates of per capita consumptions imputed into the 2011 Census for the 1,291 GN Divisions.

We investigate five main questions: 1) To what extent can variation in village economic well-being -- poverty rates defined at the 10 and 40th percentiles of national income and average GN consumption -- be explained by high spatial-resolution features? 2) Do these features predict equally well in urban and rural villages and within different portions of the village welfare distribution? 3) Which features are most strongly correlated with these measures of well-being? 4)

Can these models predict into geographically adjacent areas? and 5) Are predictions robust to the use of a smaller sample of training data?

We find that: i) satellite features are highly predictive of economic well-being and explain about sixty percent of the variation in both village average consumption and estimated poverty headcount rates; ii) Predictions are more accurate in rural areas than urban areas, and accuracy declines only slightly when considering only poor villages. iii) Measures of built-up area and roof type strongly correlate with welfare. Car counts and building height are strong correlates in urban areas, while the share of paved roads and agricultural type are strong correlates in rural areas; iv) Out-of-sample predictions are less accurate but tend to preserve rank; and v) Reducing the size of the training data by a factor of 16 has little impact on the accuracy of the prediction.

This paper contributes to a growing literature exploring how remotely sensed data may be used to assess welfare. Traditionally, the most popular remotely sensed measure for economic applications has been night-time lights (NTL), which measures the intensity of light captured passively by satellite. Strong correlations between NTL and GDP appear at the country level (Henderson et al., 2009, Pinkovskiy and Sala-I-Martin, 2016) although within a country NTL appears more strongly correlated with density than welfare. The relationship between lights and wages or other measures of income appears weak (Mellander et al., 2013), casting doubt on its reliability as a proxy for small area estimates of welfare. Additionally, NTL is ill-suited for identifying variation in welfare within small areas because of its low spatial resolution. Even the most advanced NTL satellite, the Visible Infrared Imaging Radiometer Suite VIIRS, has a spatial resolution at nadir of approximately 1.0 km^2.[5]

Daytime imagery has recently emerged as a practical source of information on welfare, in large part due to new developments in computer vision algorithms. [6] Advances in Deep Learning such as Convolutional Neural Networks (CNN) have the capability to algorithmically classifying objects such as cars, building area, roads, crops and roof type (Krizhevsky, Sutskever, and Hinton, 2012). These objects may be more strongly correlated with local income and wealth than NTL. Furthermore, textural and spectral algorithms provide a simpler alternative to analyzing HSRI that does not rely on object classification (Graesser et al. 2012, Engstrom et al. 2015, Sandborn and Engstrom 2016). In this approach, the spatial and spectral variations in imagery are calculated over a neighborhood of pixels to characterize the local scale spatial pattern of the objects observed in the imagery. These measures, which we refer to as "texture" or "spectral" measures, capture information about an area that may not be clear from object recognition alone.

This paper also contributes to a literature exploring how supervised learning techniques from machine learning may be applied to unstructured data to reveal information about human welfare (Athey, 2017). Glaeser, Kominers, Luca, and Naik (2015) apply texture-based machine vision classification to images that are captured from Google Street View, trained using subjective ratings of the images on the basis of the perceived safety. They estimate a support vector machine model and show the fitted model can reliably predict block level income in New York City

[5] Pixel size can vary depending on the angle of the satellite relative to the ground site.

Jean et al. (2016) employs an innovative transfer learning approach, in which a set of 4,096 unstructured features are extracted from the penultimate layer of a convolutional neural network that uses Google Earth daytime imagery to predict the luminosity of NTL. These 4,096 features are then used to predict the average per capita consumption of enumeration areas (villages), taken from living standard measurement surveys using ridge regression to prevent overfitting. The resulting model predicts well and explains an average of 46 percent of the variation in village per capita consumption, out of sample, across the four countries it was trained in.

While this innovative use of daytime imagery substantially improves on the use of night time lights alone, it is not necessarily optimal for predicting poverty rates. When the top two quintiles are excluded from their sample, restricting the sample to those below twice the international poverty line, the r^2 falls precipitously, to about 0.12. This illustrates the challenges this method faces in in distinguishing welfare among the poorest of the poor, who in the African context most likely live in relatively dark.. It also raises questions about the ability of a transfer learning method calibrated to night time lights to accurately predict variation in local poverty headcount rates, the measure of poverty typically of greatest interest to policymakers and aid organizations for targeting purposes.[8]

This study utilizes imagery features that are based either on recognizable objects or "texture" algorithms developed for computer vision applications, derived from High Spatial Resolution Imagery (HSRI). This method offers several advantages for the estimation of poverty rates. First, it eliminates reliance on NTL, which is a coarse measure of welfare, to identify relevant features for model development. Second, it provides a more transparent understanding of the underlying factors that explain geographic variation in welfare in different contexts. Third, features developed from HSRI, such as roads and the extent of built-up area, are useful for policy analysis in other areas as transport and urban planning. Finally, a feature-based approach can easily be extended to alternative welfare indicators, such as headcount poverty rates measured at different thresholds.

The main drawback of this feature based approach at this time is scalability. Selected features used in the study, such as roof type, road type and road width, have not yet been fully automated using existing machine learning algorithm. The rapid pace of improved computing power and algorithm development, however, mitigates this concern. Furthermore, models that exclusively use fully-automated features, such as built-up area, cars, shadows, and a full set of texture features, maintain strong predictive power. Overall, the model's strong performance illustrate the benefits of convenient access to a variety of features derived from very high spatial resolution imagery for small-area poverty estimation and program targeting.

The paper proceeds as follows: Section 2 summarizes how the data were created and presents brief summary statistics. Section 3 presents the statistical methodology. Section 4 examines the predictive power of high resolution satellite features (HRSF) to estimate poverty in small areas at the village level. Section 5 examines out of sample performance using two applications from estimating local area economic well-being. Section 7 concludes.

[8] It is not straightforward to generate estimates of poverty headcount rates from predictions of mean consumption at the village level, since poverty rates depend on the dispersion of welfare within each village as well as its mean.

2 Data Description

The analysis is restricted to a sample area of approximately 3,500 km^2 in Sri Lanka, shown in figure 1 highlighted in white. We purchased HSRI covering 47 Divisional Secretariat (DS) Divisions, the administrative area one level higher than GNs. National coverage was not feasible due to the high cost and partial availability of high-resolution imagery. We sampled DS Divisions conditional on HSRI being available, drawing equal areas from urban, rural, and estate sectors.[9] The total sample contains 1,291 GN divisions, with each GN division covering an average area of 2.12 square kilometers.

Figure 1: Coverage Area of High Resolution Satellite Imagery

Notes: Sample area shown highlighted in white.

2.1 Derivation of GN Welfare Statistics

Ideally village poverty and consumption statistics would be generated directly from the 2012/13 Household Income and Expenditure Survey (HIES), a detailed survey that measures the consumption patterns of 25,000 households on approximately 400 items. The survey only contains an average of 8.4 households per GN Division in the 47 sampled DS Divisions, making the HIES insufficient to generate consistent poverty estimates at the GN Division without supplementary

[9] Sri Lanka is unique in that it classifies sectors as urban, rural, or estate. The Estate sector is a classifications by the Department of Census that refer to "plantation areas, which are more than 20 acres in extent and have 10 or more residential laborers." For the purposes of this study, except for sample stratification, the estate sector is grouped together with the rural sector.

data. We therefore draw on the method described in Elbers, Lanjouw, and Lanjouw (2003) to impute welfare estimates into the 2011 Census of Population and Housing, which is identical to the method used to generate official poverty estimates at the DS Division level (Department of Census and Statistics and World Bank, 2015). [10] For each household in the census, per capita consumption was estimated based on models developed from the HIES, using household indicators that are common to both the Census and the HIES.[11]

Imputing welfare into the census requires an assumption of spatial homogeneity within small areas. This assumption "may severely underestimate the variance of the error in predicting welfare estimates at the local level in the likely presence of small-area heterogeneity in the conditional distribution of expenditure or income." (Tarozzi and Deaton, 2009). To test the extent of spatial heterogeneity in practice, small area estimates of poverty have been compared to census-based measures in Mexico and Brazil, which each collect income information in their census. Considerable spatial heterogeneity is present in Mexico.[12] In contrast, Elbers et al (2009) finds significantly less in Minas Gerais, Brazil. The effect of spatial heterogeneity on the results presented below is unclear. We are not aware of any empirical estimate of the extent to which spatial heterogeneity assumption leads to biased poverty headcount estimates at the local level. To the extent any additional noise in the poverty estimates due to uncaptured heterogeneity in the coefficients is independent across neighboring households within a GN, this noise would be significantly reduced after averaging over a large number of households. In the sample of 1201 GN Divisions considered below, there are an average of 2,400 households per GN.

The presence of small area poverty estimates for Sri Lanka allows for the prediction of poverty measures in addition to average per capita consumption. We derive village headcount poverty rates using the standard measure (Foster et al., 1984), for two poverty lines -- -- one at the 10% and a second at 40% of the national distribution of per capita consumption.[14] Figure 2 plots the histograms of poverty and consumption. The lower relative poverty rate (10%) shows considerable skewness. Many villages have zero estimated poverty at this lower threshold, shown as a mass point near zero. At the higher relative poverty threshold, the number of villages with zero estimated poverty is lower and the distribution of poverty rates is closer to normal. Average village consumption (income) in log points is present in the bottom center panel of figure 2, showing a distribution that is possibly bimodal and left skewed.

[10] The term welfare is used interchangeably with per capita household consumption.
[11] Consumption aggregates have been spatially deflated using a district level food price index constructed from unit values in the HIES survey by the Department of Census and Statistics.
[12] Simulations indicate that in 10 percent of municipalities, the coverage rate of the estimated poverty rate is less than 50 percent. In other words, in these 10 percent of municipalities, confidence intervals from simulations that estimate headcount rates exclude the true poverty rate in more than half the simulations.
[14] According to the official poverty line, which was developed in 2002 and subsequently updated for inflation, 6.7 percent of the population was poor in 2012/13.

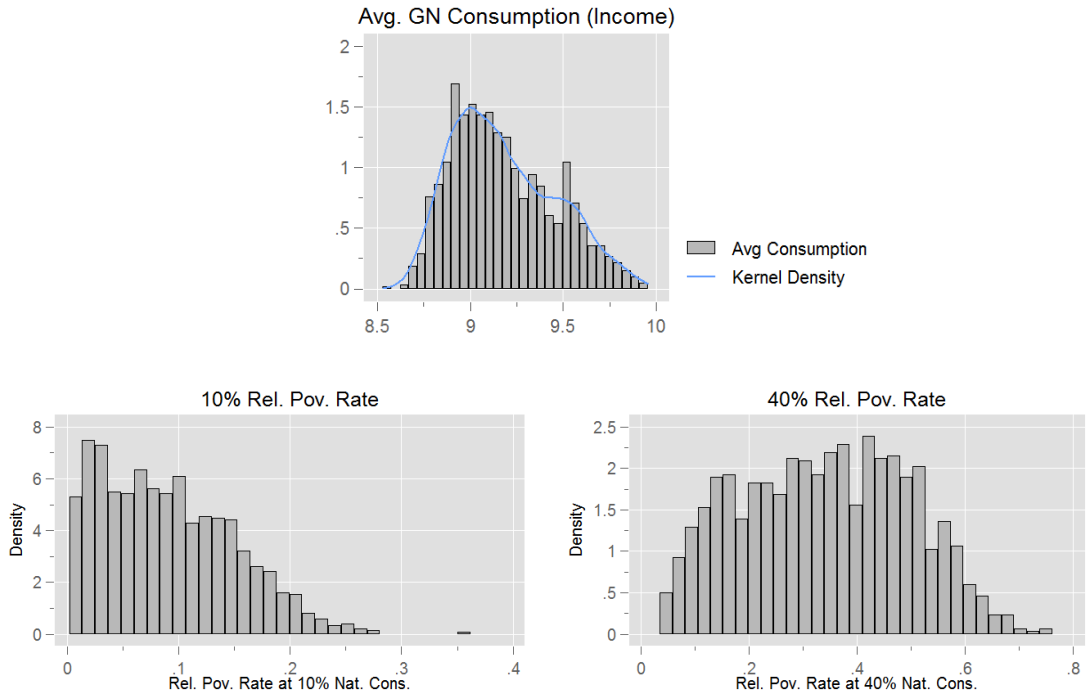

Figure 2: Distribution of Poverty and Average Log GN Consumption

Notes: Histogram of village relative poverty. The two relative poverty rates refer to the fraction of individuals in each village below 10% or 40% of national income. Data is sourced from the poverty estimates imputed into the 2011 Sri Lankan Census.

2.2 Comparison of GN Poverty Rates and Mean NTL Reflectance

A simple visual comparison between mean NTL and village poverty rates illustrates why NTL provides limited information on sub-national welfare. Figure 3 presents a panel of three images for the Divisional Secretariat of Seethawaka: mean raw NTL (left), poverty rates derived from the 10% national income threshold (middle), and log of mean population density (right). Comparing the left and middle panels, there is only a small association between villages that have low NTL reflectance and those that are high in poverty. Problems of overglow (Henderson et al., 2012) mean that poor villages adjacent to wealthy ones will be misclassified as non-poor. While NTL tracks the general contours of poverty for the DS – lower poverty areas in the Northwest and higher poverty areas in the Southeast – this coarse association is only of limited use for public policy applications such as poverty targeting or budget allocations.

NTL appears to give a more accurate approximation of the population density of the underlying GN Divisions, which is consistent with Mellander et al. (2013). Comparing the right and left panel shows a strong association between high NTL areas and areas with a high population density. We take this to suggest that the information content contained within NTL related to human welfare is limited. While lights at night may indicate gross associations, it is highly imperfect measure of

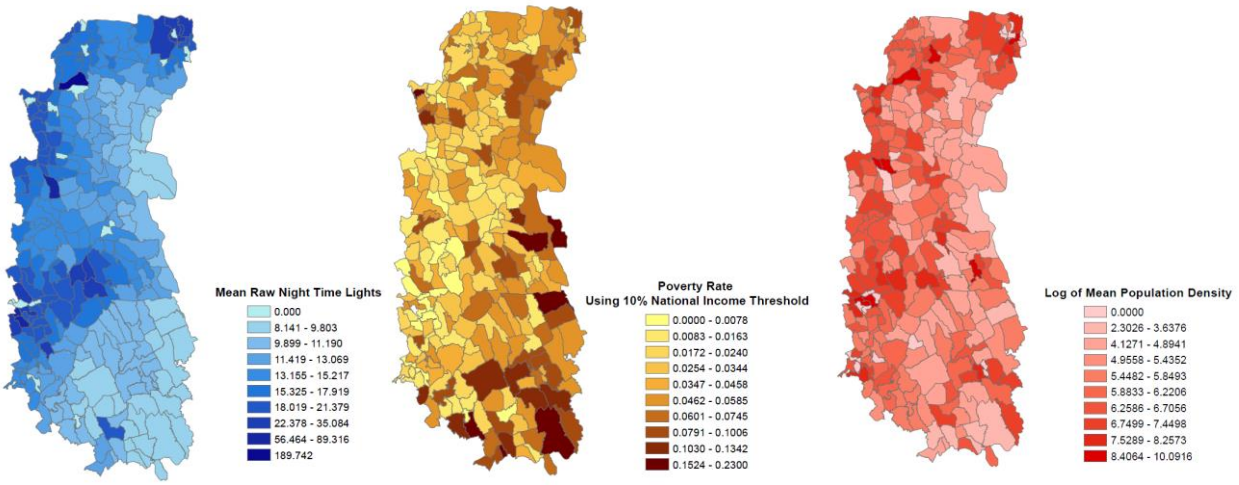

Figure 3: Comparison of Mean Night Time Lights (NTL), Poverty Rate, and Mean Population Density, Seethawaka, Sri Lanka

welfare. We therefore investigate whether the much richer set of information contained in HSRI daytime imagery translates into more accurate welfare predictions.

2.3 Description of High Resolution Spatial Features (HRSF)

The derived high resolution spatial features fall into seven broad categories: (1) Agricultural Land, (2) Cars, (3) Building Density and Vegetation, (4) Shadows (5) Road and Transportation; (6) Roof Type; and (7) Textural and Spectral characteristics. In addition to the satellite features, we use two geographic attributes of the GN Division: Whether it is administratively classified as an urban area, and its area in square kilometers. Table 1 presents summary statistics for these variables. The agricultural land indicators are coarse, and consist only of the fraction of GN agriculture identified as paddy (rice cultivation) or plantation (cash crops such as tea). These sum to one hundred percent for GNs with agricultural land, so the excluded category is GN Divisions with no agricultural land. Agriculture in the sample is evenly split between paddy and plantation. We also calculated the fraction of total GN area that is either paddy, plantation, or any agriculture.

Three car related variables were calculated – the log total number of cars in a GN, total cars divided by total road length, and cars per square kilometer of the GN. The average GN Division in our sample contains 50 cars. However, there is wide dispersion, as the 99th percentile of the car count distribution is equal to 577 cars and the maximum value is 4,000 cars. On the left side of the distribution, 136 out of 1291 GNs contain no cars. Because the distribution is skewed, we take the log of the car count, while imposing a smooth function for GNs with zero or few cars [15]

Building density variables include the fraction of an area covered by built-up area and the number of roofs identified, Built up area captures any human settlements – buildings, homes, etc. – regardless of use or condition. These are grouped with two measures of the Normalized Difference Vegetation Index. Although technically a spectral characteristic, the presence of vegetation in

[15] The log car variable is calculated as the log of the sum of the car count and the square root of the car count plus one.

urban areas indicates development such as parks, trees, or lawns (i.e., are that is not built up) within the urban environment. In the rural environment it also indicates undeveloped areas, and the values can aid in describing variations in agricultural type and productivity depending on the timing of the image acquisition.

The fifth category are two indicators that capture shadows: the log of the number of pixels classified as shadow as well as the fraction of shadows covering a valid area. The shadow variables use the angle of the sun as it shines on a building or tree, and the shadows it displaces, to estimate the presence of shadows.[16] Shadows can be caused either by tall buildings in urban areas, or by trees or vegetation in rural areas.

The road variables we calculate are the log of total road length, fraction of roads that are paved, and length of airport runway and length of railroad identified. For roof type, we calculate the fraction of roofs in a village that are either clay, aluminum, asbestos, with the omitted category being roofs that are identified as none of the above, the vast majority being gray cement roofs. Roof type can be identified through remote sensing by using hyperspectral imaging, or using reflectance from several contiguous spectral bands. Different roof materials exhibit different spectral properties, particularly in the sub-visible bands of the spectrum. The roofs in our sample are clay (36.5%) aluminum (14.08%), asbestos (7.8%) or gray concrete (41.6%).

Finally, we calculate seven separate types of spectral and textural features: Fourier transform, Gabor filter, Histogram of Oriented Gradients (HoG), Line support regions (LSR), Pantex, and Speed-Up Robustness Features (SURF). These features can be considered outputs from some dimension reduction technique, in that they are reduced dimensionality descriptions of a complex 2-D satellite imagery. These are often used in machine vision problems to decompose an image. Like dimension reduction outputs, we may have to squint at the output to determine what these spectral or textural features are capturing. They are intended to capture aspects of a neighborhood that are not so easily identified directly, including the presence of characteristics associated with slums such as many irregular building lines or high density.

Because these measures may be novel to readers without backgrounds in remote sensing, further description may be helpful. We consider Pantex here to be a measure of human settlements. It's a spatial similarity index, where each cell is compared to adjacent cells in all directions. Forests will have a low Pantex level, since cells in all directions have similar contrast, as will cells with straight roads. Cities dense with many buildings will have high Pantex values. HOG captures "local intensity gradients or edge directions" (Dalal and Triggs, 2005) and in context here captures intensity of lines of development or agriculture. Local binary patterns (LBPM) captures local spatial patterns and gray scale contrast. SURF detects local features used for characterizing grid patterns, and measures orderliness of building development, the opposite of which is typically referred to as a slum. Areas with right angles, corners, or areas with regular grid patterns, will have larger SURF values relative to areas with chaotic or irregular spacing.

[16] Valid area refers to areas at the foot of building where shadows may appear.

For more detail on imagery and the feature extraction process we refer the reader to appendix A, In brief, object-based features were classified through a combination of convolutional neural network (CNN) training and object based image assessment (OBIA). Accuracy varied by feature, but overall class accuracy was above 90% for all features[17]. To illustrate the classification process, we will discuss two training examples for built-up area (building footprint) and cars, shown in figures 4 and 5. Figure 4 shows raw satellite imagery in the right panel, and classified imagery according to the CNN algorithm in the left panel. Areas highlighted in green in are true positive building classifications, where areas classified by the algorithm as buildings were confirmed as such by manual identification. Areas in red are false positives that were erroneously classified by the algorithm as buildings even though no buildings are present. Figure 5 shows raw satellite imagery from Colombo with cars identified by the CNN algorithm are highlighted by blue circles. The classifications occur on the road paths or in parking lots where cars are expected to appear. There are some false negatives – cars in the image not classified as such – particularly in areas where cars are obscured by trees or vegetation. However, the classifier appears accurate enough to distinguish high car areas from low car areas.

Figure 4: Example Developed Area (Buildings) Classification

Notes: above image shows raw (left) and classified (right) for developed area building classifier from raw satellite imagery. Areas in green show are true positive building classifications. Images in red are false positives: erroneously classified areas as buildings.

[17] See for example the ROC curve for the buildings in figure A1.

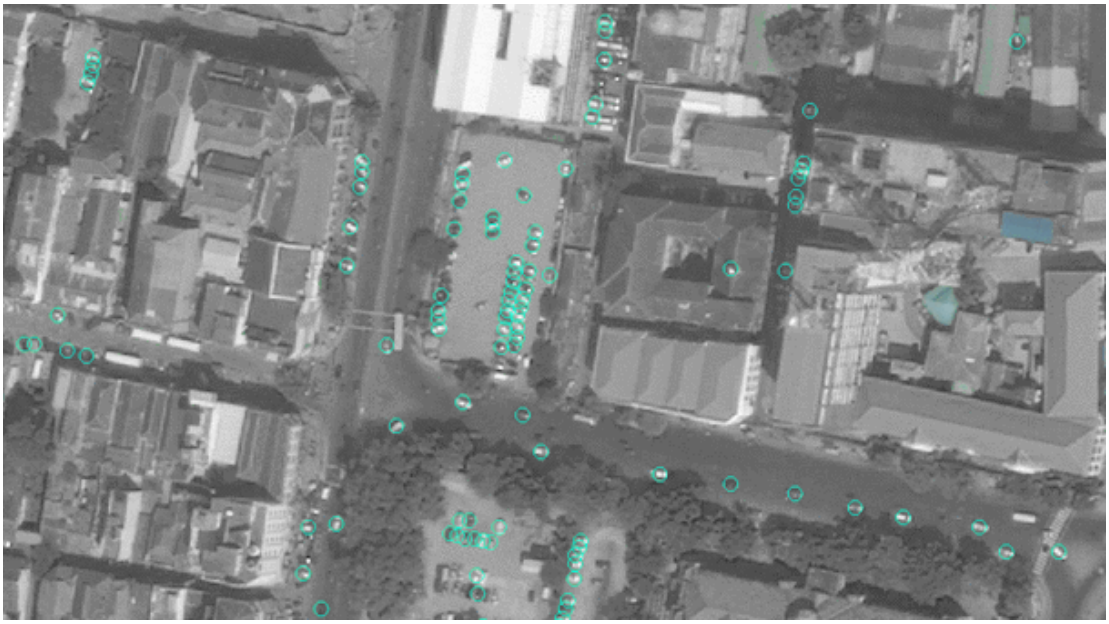

Figure 5: Example Car Classification
Notes: above image shows satellite imagery overlaid with cars identified algorithmically shown in blue.

3 Statistical Methodology

The exercise aims to validate the accuracy of poverty predictions derived from high-resolution spatial features, in the absence of additional household indicators, and establish which features contribute explanatory power. For the baseline, we report results from Ordinary Least Squares (OLS) regressions, in which we could model the poverty rate or the log per capita consumption of each GN. Given that the poverty predictions are not constrained to lie between zero and one, however, the annex reports results from a binomial regression on poverty rates.

3.1 Model Selection

Given the list of covariates in table one, variable choice is not straightforward. Estimating a model with the full set of variables in table one would likely produce predictions that are overfit, in the sense that they perform much better in-sample than out-of-sample (Athey and Imbens, 2015). One attractive method for variable selection among a large selection of covariates is Lasso regularization. Lasso is a regularized regression that estimates a regression model with an added constraint that enforces parsimony (Tibshirani, 1996). The motivation for the shrinkage estimator is that by reducing the parameters of the model, one increases bias at the expense of lower variance.

We estimate a standard lasso model

$$(1)\quad \beta_{Lasso} = \arg\min_{\beta} \sum_{i=1}^{N} Y_i - X_i\beta + \lambda \sum_{j=1}^{k} |\beta_j|$$

Where the poverty rate in a village is given by Y_i and $\lambda \geq 0$ is a shrinkage parameter that penalizes the absolute values of the coefficients. At the extreme, fully relaxing the penalization factor and setting λ to zero yields unconstrained estimates: $\beta_{BinLasso} \rightarrow \beta_{Bols}$. As $\lambda \rightarrow \infty$, the penalty increase and $\beta_{BinLasso}$ converges to the zero vector. Lasso regressions are useful as a variable selection methodology because the sharp L_1 metric leads variables to drop out abruptly as λ increases, creating a de-facto variable selection mechanism in addition to "shrinking" coefficients in magnitude towards zero (Varian, 2014). To choose the appropriate value of λ, we apply 10-fold cross validation, and choose the value of λ that minimizes root-mean squared error (RMSE) across folds. For robustness, we also estimate a binomial version of the model, which ensures that predicted values lie in between zero and one.

Inferential standard errors are typically absent from Lasso models. Because of the Oracle property of the Lasso estimator (Fan and Li, 2001), we can employ a two step "Post-Lasso" methodology: first estimate a Lasso model over the full candidate set of covariates, x, resulting in $\hat{\beta}_{BinLasso}$ and $x_{selected} \subset x$, the "selected" set of variables from the Lasso estimation where selected indicates Lasso returns a non-zero coefficient. We can then estimate an unconstrained model using only the reduced set of selected coefficients. The Oracle property ensures that inference in the second stage using the reduced set of variables selected in the first stage is consistent with inference were we to use a single stage estimation strategy using only the selected variables present in the true data-generating process (Belloni and Chernuzhukov, 2013).

4 Poverty Validation Using High Resolution Features

Table 2 presents the estimates from the main specification for the full sample. The first two columns report post-lasso coefficients on a regression of the GN Division's 10 percent and 40 percent relative poverty rate as the dependent variable, while the third column reports the coefficients for a dependent variable of average log per capita consumption.

The models generate accurate predictions of poverty and well-being. R-squared values vary from 0.608 for the average village consumption models, 0.61 for the 10% poverty line, and 0.618 for the 40% poverty line. In other words, a simple linear model that includes only the area of the GN, whether it is urban, and remotely sensed information explains 61 to 62 percent of the variation across villages in both the headcount poverty rates and economic welfare, as measured by the mean log consumption per capita of its households. Lasso regularization selects 27 out of 31 candidate variables for the 10% poverty rate models, 18 out of 31 candidate variables for the 40% poverty rate models, and 19 variables for the log per capita consumption model. T-statistics are presented and reflect clustering at the DS Division level.

While the primary objective of this exercise is to obtain accurate predictions, the model coefficients also shed light on the nature and magnitude of the conditional correlations between imagery features and poverty. The coefficients listed in Table 2, however, are difficult to interpret

for two reasons. First, the independent variables are often measured in different units, complicating comparisons between coefficients. Second, in some cases multiple independent variables are based on the same underlying feature. For example, the log number of cars, cars per road, and cars per GN are all derived from car counts. In these cases, it is meaningless to evaluate the conditional correlation of one variables while holding the others constant. Therefore, to get a sense of the relative magnitude of the conditional correlations, we group selected independent variables together and consider the marginal effect of a one standard deviation increase of all variables within the group.[19]

The results are presented in Table 3. For some dependent variables, the reported marginal effects reflect a combination of multiple underlying indicators, while for others they reflect single variables, as indicated in the right most column. The size of the GN, in square kilometers, is strongly correlated with poverty rates at the 10 percent level, much more so than 40 percent headcount or average consumption. This suggests that households in the bottom decile are disproportionately found in larger GN Divisions. The presence of agricultural land is weakly and negative associated with poverty, controlling for other characteristics of the GN, although the result is not statistically significant.

Of the indicators related to the distribution of paddy vs. plantation land, The LASSO procedure selected three of the indicators for 10 and 40 percent poverty incidence models, and two for the log consumption model.[20] The results indicates a discernible but fairly weak negative relationship between the presence of paddy agricultural land and poverty, which is consistent with the traditionally disadvantaged nature of the tea plantation sector in Sri Lanka.

Compared with land type, the association between poverty and cars is mildly stronger. The standard deviation average GN has average log car count of 3.1. A one standard deviation increase is associated with a 2.2 percentage point decline in poverty at the 40 percent threshold, and a 0.035 increase in predicted log per capita consumption. The model explaining the 10 percent poverty rate includes two additional car related variables, total cars as a share of road length and cars as a share of total area. A one standard deviation increase in all three variables, is associated with a 1.2 percentage point decline in poverty at the 10 percent rate.

Road characteristics are moderately associated with local poverty rates. The lasso algorithm selected all four of the roads variables for the 10 percent poverty and average consumption models, which are total length of road, fraction paved, the log of airport runways, and the log length of railways. The first three are negatively associated with poverty, though only the first two are statistically significant, while GNs with more railways are poorer. Total length and fraction paved are more strongly associated with poverty, and were the only two were selected for the 40 percent poverty rate model. A one standard deviation change in total length of road is associated with a 1.9 percentage point decline in 10 percent poverty, a 2.5 percentage point decline in 40 percent poverty, and a .031 increase in log consumption. The magnitudes of the marginal effects for

[19] Except for percent of GN agriculture that is plantation, for which a one standard decrease is considered.

[20] Since an increase in paddy land implies a reduction in agricultural land, for those GNs with agricultural land, the latter is subtracted instead of added when calculating the marginal effect.

fraction of roads that are of paved are broadly similar, though a one standard deviation increase is only associated with a weaker 1 percentage point decline in 10 percent poverty.

The two measures of building density, the log number of total buildings and share of area that are built-up, are strongly associated with log welfare and poverty. A one standard deviation increase in these two variables is associated with a 2.7 percentage point decline in the 10 percent poverty rate, an 8.1 percentage point decline in the 40 percent rate, and a 0.16 increase in log consumption. In the 10 percent poverty model, a one standard deviation increase is associated with a smaller 2.7 percentage point decline in poverty.

Vegetation, which is to some extent the converse of built-up area, is moderately associated with poverty. A one standard deviation reduction in vegetation is associated with a 2.9 percentage point reduction in 40 percent poverty, and a .04 increase in mean per capita consumption, which is comparable to cars or the fraction of roads that are paved For the model explaining variation in 10 percent poverty, two measures of the Normalized Difference Vegetation Index (NDVI) were selected, calculated over two different sized blocks. This reverses the positive association between vegetation and poverty in the model, meaning that a reduction in vegetation is associated with slightly increased poverty, although the marginal effect is not statistically significant. The 40 percent poverty and log welfare models only include NDVI calculated over blocks of 64 pixels, suggesting that very high spatial resolution imagery may not be critical for generating informative measures of NDVI for prediction.

Two measures of shadows are selected, the share of valid area covered by shadows, and the log number of shadow pixels.[21] As mentioned above, shadows can reflect either the presence of buildings or trees. A one standard deviation increase in both measures is associated with a 3 percentage point increase in 10 percent poverty, an 8 percentage point increase in 40 percent poverty, and a 0.13 decrease in mean log per capita consumption. By this measure, the association between shadows and poverty is strong, comparable to building density.

For roof type, the LASSO procedure selects both the fraction of roofs classified as clay and aluminum, for all three models, and includes the fraction classified as asbestos for the 10 percent poverty model. The signs on clay and aluminum in the poverty regressions are positive, suggesting that these are generally inferior compared to the omitted category of grey. This appears to be consistent an analysis in Kenya that documents that roofs with greater luminosity, like aluminum, are associated with lower levels of poverty (Suri et al., 2015). The marginal effect of a standard deviation in clay and aluminum roofs are, respectively, 1.7 and 0.6 percentage points for 10 percent poverty, and .06 and .03 for mean log per capita consumption. These magnitude are stronger than roads and vegetation, but considerably less than those for building density and shadows.

Of the texture variables, five out of seven are selected for the 10 percent model (LBPM, LSR, and Gabor, Fourier, and SURF). Of these, only LBPM and SURF are selected for the 40 percent and log per capita consumption model. In general, the estimated marginal effects for these variables are modest. The main exception is the mean of the Fourier transform, which is positively associated with 10 percent poverty, though the coefficient is not statistically significant. A one standard deviation increase in SURF is associated with a percentage point decline in the 40 percent rate and

[21] Valid area refers to all land are (excluding lakes and rivers).

a 0.03 increase in log per capita consumption. This is consistent with wealthier areas being laid out in a more orderly way, with more "right angles" in housing layouts. More research is needed to better understand how these spatial features may be capturing building density or other visual characteristics associated with poverty.

As a further check the accuracy of the model, Figure 6 plots the predicted economic well-being against true economic well-being in figure 6.[23] Each point represents a village, where the location on the x-axis corresponds to the true poverty rate and location on the y-axis is the model predicted poverty rates. A model that is perfectly able to predict poverty using satellite variables would represent a 45 degree line, starting at the origin and ending at the upper right of the graph. For all of the measure of economic well-being, the points on the graphs roughly straddle the 45 degree line. The predictions are clearly able to distinguish high poverty areas from low poverty ones, albeit with inevitable noise.

Finally, figure 7 presents a map showing the true welfare measures on the left panel, against the predicted welfare measures on the right, for a particular DS Division, Seethawaka. The top panel shows predicted welfare from the OLS model against actual welfare. The model is able to distinguish the poorer eastern areas from the richer western ones. Even poor GNs adjacent to richer ones can be distinguished; although the smallest GNs are less than a half mile across, the HRSF model is able to distinguish with considerable accuracy the variation in average consumption. The middle panel shows predicted and true poverty rates defined at the 10% relative poverty level. Again, the predicted model approximates the true poverty rates with considerable accuracy. The lower poverty regions in the south and north east are replicated in the predicted values. The model tends to under-predict poverty in the lowest poverty areas in the mid-west, suggesting that two step or zero-inflated Poisson models, which are commonly used to model distributions with many zeros, may perform better.

In sum, predictive models based on an urban indicator, the size of the GN, and a host of features derived from satellite imagery predict poverty rates and mean log per capita consumption remarkably well. In general, the size and magnitude of the coefficients are sensible. Larger and rural villages tend to be poorer, as are villages with more agricultural land devoted to plantation rather than paddy crops. Greater numbers of cars are is associated with lower poverty, although the relationship is not statistically significant, as is a denser road network and a larger share of paved roads. The indicators most strongly associated with poverty are building density and shadows. Shadows are positively associated with poverty, which suggests they are capturing variation in tree cover that is inversely related to building density. Consistent with this, areas characterized by more and lusher vegetation tend to be poorer. Clay and aluminum roofs, compared to grey roofs, are associated with greater levels of poverty. Of the spatial features, SURF exhibits a fairly strong association with poverty at the 10 percent rate, suggesting that neighborhoods laid out in a more orderly way tend to be less poor. The following sections consider the robustness of these main findings.

[23] "True economic well-being", as noted above, is the mean of predicted per capita consumption in the census.

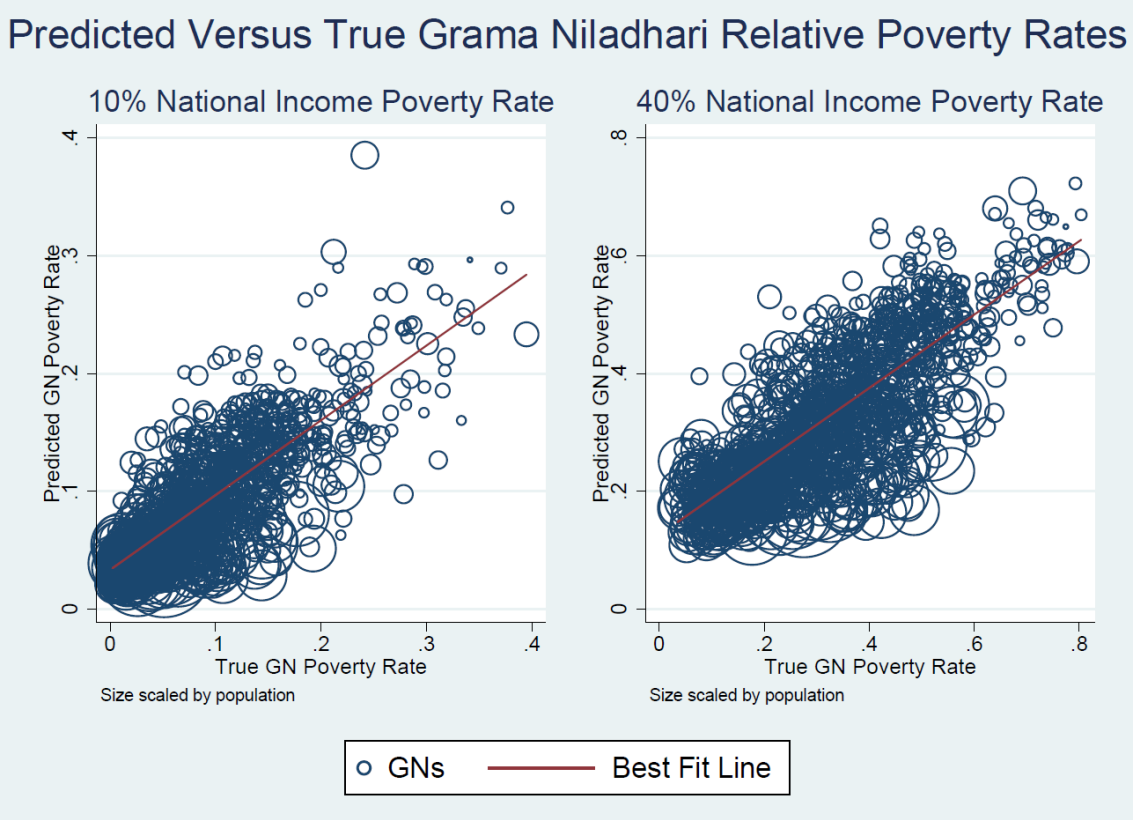

Figure 6: Predicted Versus True Welfare Measures, 10% Relative Poverty Rate (left), 40% Relative Poverty Rate (right)

4.1 What share of the explained variance is accounted for by different features?

The results presented above indicate that features derived from satellite imagery explain a large portion of village income or poverty, and that associations are particularly strong for measures of building density and shadows. However, these results don't address the question of which indicators account for the model's predictive power. To address this issue, we decompose the R^2 using a Shapley decomposition (Shorrocks, 2013; Huettner and Sunder, 2012; Israeli, 2007). This procedure calculates the marginal R^2 of a set of explanatory variables, as the amount by which R^2 declines when removing that set from the set of variables. In other words, for a model with k sets of explanatory variables, the procedure will estimate 2^{k-1} models and average the marginal R^2 obtained for each set of independent variables across all estimated models. This ensures that the variable's contribution to R^2 is independent of the order in which it appears in the model.

Table 4 presents the R^2 decomposition. The results confirm that measures of building density – built up area, number of buildings, shadow pixels, and to a lesser extent vegetation, are powerful contributors to predictive power. Collectively, these three sets of variables account for 37 to 40 percent of the model's explanatory power. However, a number of variables are moderately important. GN area, urban classification, Road characteristics, roof type, and the texture

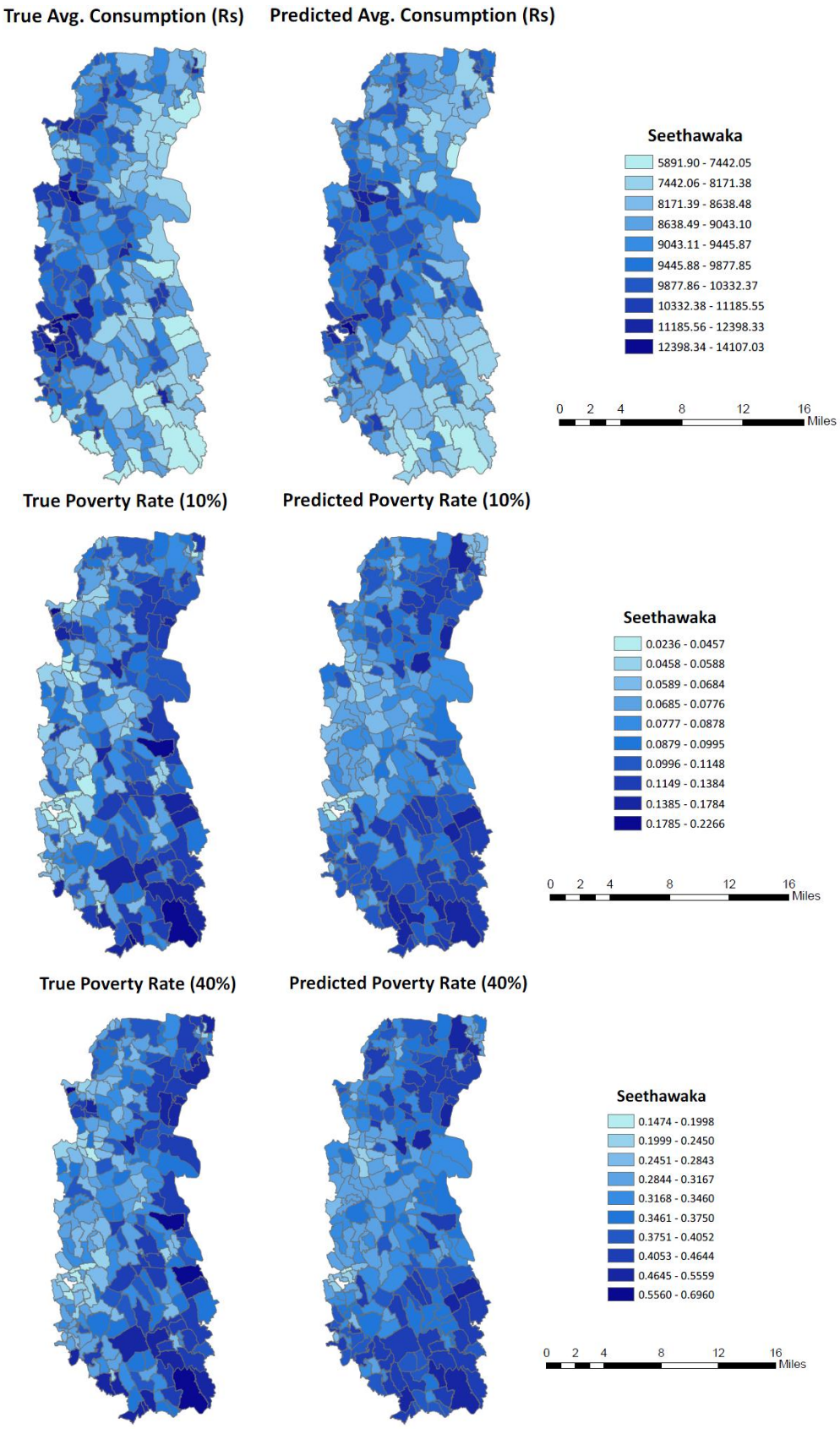

Figure 7: Predicted Versus True Welfare Measures, Binomial Logit Models, Average Consumption (top), 10% Poverty (middle) 40% Poverty (bottom)

variables each explain 8 to 12 percent of the variation. The car and agricultural variables explain a bit less than that, between 5 and 7 percent each. In short, while broad measures of building density explain a large share of the variation, virtually all sets of indicators contribute substantial predictive power to the model.

4.2 Comparisons to Night Time Lights

How do the predictive power of indicators derived from daytime imagery compare with night time lights? To shed light on this, Table 5 presents OLS models covering the same sample area using night time lights as the independent variable. The first three columns present poverty and per capita consumption models. Aggregate night time lights is positively correlated with welfare and negatively correlated with poverty, however the total explanatory power is low: R^2 values for the three regressions are between 0.10 and 0.147, with performance lowest for the 10 percent headcount measure and highest for log consumption per capita. Adding higher order polynomials up to a quartic only increases it to 0.15. Models built using high resolution satellite indicators capture around four times as much variation in poverty or welfare as NTL. Columns 4-6 of table 4 shows estimates that include DS Division fixed effects. Night time lights is no longer significant in any of the specifications, indicating that within DS Divisions, NTL is weakly correlated with welfare.

Given the prevalence, ease of use and familiarity with night time lights, one might also ask how much more explanatory power do night time lights provide *in addition to* the indicators extracted from daytime imagery? Table 6 answers that question, by adding night time lights to the Shapley decomposition reported in Table 4. The night time lights category includes average, squared, cubed, and average standard deviation of NTL. The night time lights variables explain between 7 and 12 percent of the variance in per capita consumption or poverty according to the decomposition, meaning there is roughly a 90 percent additional variation in poverty or income that is captured through high resolution satellite predictions. Furthermore, adding night time lights marginally increases the overall r^2 of the regression, by about 0.01. In this context, NTL is not a particularly accurate proxy for poverty and welfare, and adds very little explanatory power to the set of available daytime indicators.

4.3 Urban and Rural Linear Models

How does the relationship between indicators and welfare differ in urban and rural areas? Table 7 shows model estimates estimated separately for 393 urban villages and the 898 rural ones, based on Sri Lanka's prevailing definition of urban and rural areas.[24] Variables were again selected through Lasso estimation. The urban model selects fewer variables – 13 of the candidate variables in the urban model are selected versus 16 for the rural model. R-squared values are slightly higher in rural areas (0.656) and significantly lower in urban areas (0.445).[25] For the urban model, log number of cars, built-up development, and shadow pixels are important. In rural models, agricultural variables, roof type, shadow pixels, NDVI, Pantex and LBPM are important. Two

[24] This definition is based on administrative units and has not been updated in many years. As a result, some areas officially classified as rural have urban characteristics.

[25] This might be due to the presence of de-facto urban GNs in the rural sample. In addition, the nature of the consumption module in the HIES, which could better capture consumption in rural than urban areas.

particular coefficients make the results more plausible. First, the association between cars and poverty is significantly stronger in urban areas. Second, the association between NDVI and poverty is strongly negative in rural areas, as rural areas with more vegetation and less built-up area are poorer. The coefficient on NDVI in urban areas, meanwhile, is positive and not statistically significant, suggesting that if anything wealthier urban GNs are characterized by a greater prevalence of lush vegetation.

4.4 How accurate are models predicting log per capita consumption for different portions of the welfare distribution?

The model's ability to predict variation in headcount poverty rates at two thresholds suggests that it can effectively distinguish between households within lower parts of the welfare distribution. To verify this, we divide the sample of GN Divisions into quintiles based on the mean predicted per capita consumption of census households, and re-estimate the main model for log per capita consumption on the subsample of the bottom 80, 60, 40, and 20 percent of the distribution. Model performance statistics are shown in Table 8. Overall, the model continues to predict well within poor subsamples, as the adjusted r-squared declines only mildly from 0.6 in the full sample to 0.52 when only considering the bottom quintile. Mean absolute error varies from 0.06 in the bottom quintile to 0.14 in the full sample, both of which are about 0.5 standard deviations of the dependent variable. The coefficients are not reported, but within the bottom quintile, those on area, the percent of land devoted to agriculture, NDVI, Cars, and SURF remain statistically significant despite the large reduction in the number of observations. Coefficient estimates on those variables are generally similar within the bottom quintile, although the relationship between SURF and poverty is negative.

4.5 Correcting for Spatial Autoregression

One unaddressed concern is whether the presence of either spatial autocorrelation or spatial heterogeneity leads the standard errors to be underestimated. Spatial autocorrelation can occur in the presence of geographic spillovers or interactions (Anselin, 2013), and considering the village-level observations one could develop plausible stories by which poverty is influenced by this mechanism. A Moran's I test for the presence of such disturbances according to Anselin (1996) rejects the null hypothesis that there is no spatial autocorrelation present.

To correct for the spatial autocorrelation we model explicitly the spatial autoregression (SAR) process and allow for SAR disturbances, a so called SARAR model. This is implemented via a generalized spatial two-stage least-squares (GS2SLS) as shown in Drukker et al. (2013). The results presented in table 9 show that after correcting for spatial autocorrelation most high-resolution spatial features remains significant predictors of local area poverty. Although there is some presence of autocorrelation, it is not sufficient to alter the joint significance of the spatial variables.

4.6 Do High Resolution Satellite Features Explain the Poverty Gap?

The poverty gap is a useful supplement to the headcount rate for understanding poverty because it takes the depth of poverty into account. The poverty gap or FGT_1 metric measures poverty

depth by considering how far the poor are from a given poverty line. [27] We compute the average poverty hap for each village, and use this measure as a dependent variable in a regression where the right hand side includes the size of the GN, a dummy indicating urban classification, and the features created from high resolution satellite imagery. As in the headcount rate, we consider poverty lines at both the 10^{th} and 40^{th} percentiles of national per capita consumption. Table 10 presents the results estimated via OLS. The coefficients can be interpreted as a unit change in the distance between the poverty gap and the poverty line for the average village. As was the case for headcount rates, high resolution features explain the poverty gap well, with adjusted R^2 values between 0.588 and 0.609. Not surprisingly, building density and shadow variables are also strong correlates of the poverty gap.

5 Out of Sample Performance with Two Applications

5.1 Poverty Mapping Using Partial Census Sample Size Combined with HRSF

The prevailing method for estimating poverty in local areas is to impute household welfare into a full census using a nearly contemporaneous welfare survey. A key motivation for the analysis above is to assess the extent to which satellite imagery is a viable substitute for census data. In other words, can a household survey be combined with imagery to produce sufficiently precise small area estimates? To assess this, we examine whether the predictive power of satellite imagery remains when it is calibrated using a census extract rather than a full census. If predictions of GN poverty based on satellite indicators remain accurate when estimated using a sample extract, then a sufficiently large household survey can be combined with imagery to produce small area estimates at a fraction of the cost of collecting a full census.

For this exercise we produce several simulations of the dependent variable (either per capita consumption or poverty rate) using a Census with reduced sample size. We compute subsamples of 25% and 50% of villages (GNs) on which to train our high resolution models[28]. We also vary the number of households the sample and subsample that are "surveyed". Welfare or poverty is measured in each GN using either 25%, 50% or 100% of the actual households in that GN. For example, a 50% GN sample where 25% of the households are surveyed estimates a model using poverty rates calculated from half of the villages, and a quarter of the households within each village. The estimated actual poverty rate of a village will become less precise the fewer households that are sampled per village. Similarly, reducing the number of villages in the sample makes the estimated coefficients of the model less precise. Reducing the number of villages in the sample saves more money than reducing the number of households sampled per village. This exercise is intended to illuminate the resulting cost, along both margins, in terms of the accuracy of the prediction.

Table 11 presents model performance in this simulation exercise. We present in-sample and out-of-sample R^2, which calculates the coefficient of variation for villages included or excluded from

[27] We calculate for our sample the FGT_1 metric (Foster Greer and Thorbecke, 1984), which is defined as FGT_1=1/N ∑_(i=1)▒((z-y_j)/z) , where y_j is an individual's income, and z is the poverty threshold.

[28] Villages enter the training or test sets randomly, which differs from the exercise in the following section in which villages enter or exit the training set systematically based on contiguous geography.

the training sample. First, out of sample R^2 is slightly *higher* than in-sample, indicating the models in the previous section may if anything be underfit. For the 10% relative poverty rate models, out of sample R^2 varies between 0.637 and 0.621 while in-sample R^2 lies between 0.605 and 0.595. Estimated out of sample R^2 increases only slightly as the number of sample villages (GN Divisions) increase. In terms of the average error rate of individual predictions, we present normalized mean absolute error, which computes mean absolute error expressed as a percentage of average poverty rate or income. Average error increases marginally when fewer households per village are sampled. Average normalized error is around 1/3 of the poverty rate for the 10% poverty level, ¼ of the poverty rate for the 40% poverty measure, and one to two percentage points for average village welfare, regardless of number of villages or households sampled per village. These results suggest that HRSF can act as a substitute for census data when generating local estimates, possibly making it cost-effective to generate local poverty estimates in between censuses.

5.2 Poverty Extrapolation to Adjacent Areas Using HRSF Models

A strong motivation for using satellite imagery is to extrapolate poverty estimates into areas where survey data on economic well-being does not exist. While most of the data deprivation that characterizes the developing world occurs at the country level, it is also common for surveys to omit selected regions, due to political turmoil, violence, animosity towards the central government, or prohibitive expense. For example, from 2002 through 2009/10, Sri Lanka's Household Income and Expenditure Survey failed to cover certain districts in the North and Eastern part of the country due to civil conflict, and Pakistan's Household Income and Expenditure Surveys exclude the Federally Administered Tribal Areas and Jammu and Kashmir.

To assess how well a model "travels" to a different geographic area, we use a form of "leave-one-out cross-validation" (LOOCV), a common method used to infer statistical out of sample performance (Gentle et al., 2012). In standard LOOCV, a model is fit for every observation excluding one, then the estimated relationship is used to predict into the withheld observation. This is repeated for every observation in the dataset until every observation has an associated predicted value. This ensures that for all observations i, the fitted value of $\hat{y}_i$ used to build the model is not influenced by the relationship between x_i and y_i.

Our approach differs from the standard case in that for each estimation we exclude, or "leave out", an entire Divisional Secretariat (DS), an administrative sub-unit at the level immediately below the district. To give a sense of size, our sample contains 47 unique DS divisions. This type of LOOCV is a more stringent test of out of sample performance, but one that more accurately approximates the intended use-case of extrapolating poverty into areas where data are not present. While traditional LOOCV assesses out-of-sample performance relative to a large set of single observations, omitted at random, most cases of incomplete survey coverage omit one or more regions within a country.[29] If the data generating process for poverty is geographically heterogeneous, LOOCV at the DS level will give a more accurate assessment of our methodology in practice.

Our algorithm for adjacent prediction LOOCV is as follows:

[29] A further complication is uncovered regions are not selected at random and are likely differ from the surveyed regions in unobserved ways. This will contribute to prediction error, since extrapolation requires assuming that the model estimated in the surveyed regions applies to the uncovered region as well.

1. Estimate a prediction $\widehat{f}_{\{i\}}(x_{-i})$ on all Divisional Secretariats save for holdout DS i.
2. Use the estimated model in step 1 to predicted values for withheld DS $y_i = \widehat{f}_{\{i\}}(x_i)$.
3. Repeat until all DSes have predicted values.

To assess accuracy we use three different metrics to compare predicted poverty rates against the true poverty rates at the GN Division level. 1) Normalized root mean squared error (NRMSE) $= \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\hat{y}_j^{CV} - y_j\right)^2}$, where $\hat{y}_j^{CV}$ is the cross-validated predicted village economic well-being, y_j is the true economic well-being, and j indexes each GN. 2) Normalized mean absolute error as defined as in the previous section, and 3) Spearman rank correlation[30].

Table 11 shows the simulation results. The adjacent prediction error rates are low for the average consumption models: NRMSE is estimated at 0.083 and NMAE at 0.241. For the poverty rates, adjacent prediction error is higher. At the 10% poverty threshold NRMSE is 0.559, and NMAE is 0.404. For the 40% poverty threshold NRMSE is 0.363 and MAE is 0.276. These error rates may not be sufficient for calculating official statistics, but they may be sufficient for generating rank ordering of villages by poverty or income. The correlation between the predicted and the true values confirms this. Spearman's ρ is estimated at between 0.68 and 0.7 for the three models. We conclude from these results that high resolution imagery cannot yet be used to predict accurately into adjacent areas. However, our choice of model may also be in influencing out of sample behavior. Models with better non-linear properties, or models that can more easily handle heterogeneity – such as random forests, SVM, ensemble methods, or deep learning – may exhibit better performance when extrapolating into adjacent areas.

6 Conclusion

Traditionally, given the prohibitive cost of conducting surveys sufficiently large to provide accurate statistics for small areas, generating small area poverty estimates require pairing a welfare survey with a census or intercensus survey. Census and intercensus data is produced relatively infrequently and with a lag, making it difficult to rapidly assess changes in local living standards. The results above show that indicators derived from high spatial-resolution imagery generate accurate predictions of local level poverty and welfare, and that by and large the conditional correlations are of sensible signs and magnitudes. Furthermore, predictions based on specific features accurately predict mean per capita consumption throughout the welfare distribution. The results strongly suggest that ancillary data derived from HSRI, when paired with a household

[30] Spearman's rank correlation is a non-parametric method that measures only the correlation in the monotonic rank ordering between two variables, in this case the predicted and true poverty rates. This metric indicates whether the GN divisions can successfully be ranked on the basis of their predicted poverty rates. A rank correlation coefficient of 1 would indicate that models using only HRSF can create a perfect rank ordering of village economic well-being using the more stringent DS leave one out cross validation.

survey, can act as a viable substitute for census data in the production of local poverty estimates. While the welfare consequences of more frequent measures of poverty and inequality are unknown, they may be large given the many applications of frequent local measures of economic well-being, ranging from impact evaluation, to budget allocation to social transfers.

How well do indicators derived from satellite imagery predict poverty and which indicators are most important? We investigate these questions using a sample of 1,291 villages in Sri Lanka, linking measures of economic well-being with features derived from high resolution satellite imagery. The results indicate that the correlation between satellite derived indicators and economic well-being is remarkably strong. Simple linear models explain 60 to 61 percent in the variation in poverty and average log per capita consumption. In both rural and urban areas, variables measuring building density, built-up area, and shadows are the strongest predictors of variation in poverty.. As expected, the extent and lushness of vegetation is negatively correlated with welfare in rural areas, and mildly positively correlated with incomes in urban areas, suggesting that trees and other vegetation are a luxury in urban areas.

The analysis also included several mathematical transformations indicating the "texture" of the image, which have been utilized for optical recognition purposes and other imagery processing applications. As a whole, these are correlated with poverty at both the 10 and 40 percent rate. They generally appear to capture contrasts and sharp edges, which characterize wealthier areas, although additional analysis is needed to better understand which features explain variation in welfare, at what scale they should be calculated and what specific characteristics of the built environment they are measuring. The major advantage of these "texture" features is they provide insight into landscape variability without having to train Convolutional Neural Networks to recognize objects, as they are simply characterizing the variability in the imagery. Because of this, these variables are simple and straightforward to calculate and require no training data to create.

While these results are very encouraging, additional analysis suggests caution when extrapolating predictions into geographically adjacent areas. The normalized error rates range from a quarter to one half of the poverty rates, depending on the incidence of poverty. The likely impediment to extrapolation is geographic heterogeneity in the relationship between indicators and welfare. Another factor is time differences at which satellites images were taken, which can contribute to noise in the independent variables across geographic regions. This could impact particular indicators such as car counts, which can vary greatly according to the day of the week the imagery was obtained. Measures of agriculture also exhibit considerable seasonal variation which may also confound extrapolation to adjacent areas. This suggests that some indicators may particularly contribute to bias when extrapolating across space, and that the date of the image is an important consideration when considering spatial extrapolation using satellite-based indicators. The rapidly increasing frequency of imagery capture should mitigate these issues over time.

These findings raise a host of questions for further work. The most immediate of these is whether satellite indicators can substitute for census data in different contexts. Does the strong correlation between satellite-based indicators and economic well-being extend to income measured directly from an expenditure survey? We were unable to test this directly because of the limited number of households in the consumption survey that lived in areas covered by sample imagery. While the strength of the results reported above are encouraging, the village poverty rates used were

generated from models that explained 40 to 50 percent of the variation in measured consumption. This suggests that satellite-based indicators can at an absolute minimum explain one quarter to one third of variation in survey-based consumption, but this not be sufficiently accurate for policy-makers. Second, it is important to better understand the extent to which these results generalize to different ecological environments, such as Africa, the Middle East, and other parts of Asia. There is no guarantee that the predictive power of building density, shadows, and documented above, for example, will hold in all environments.

A second line of research could explore whether changes in satellite imagery could be used to changes in economic well-being across space and time. Poverty surveys are typically collected every three years. Therefore, the ability to "now-cast" measures of economic well-being by combining frequently updated satellite imagery with the most recent survey-based measures of poverty has great potential. More research is needed to understand how best to nowcast recent changes in local welfare. Secondly, additional research can shed light on identifying the best way to predicting into adjacent areas not covered by surveys. More flexible modeling specifications will likely improve adjacent area predictions, and some indicators may be more geographically stable, in terms of their relationship with welfare, than others. In general, the inevitable increase in the availability of imagery and feature identification algorithms, in conjunction with the encouraging results from this study, implies that satellite imagery will become an increasingly valuable tool to help governments and stakeholders better understand the spatial nature of poverty.

References

Afzal, M., Hersh, J., and Newhouse, D. (2016). "Building a better model: Variable Selection to Predict Poverty in Pakistan and Sri Lanka". Mimeo, World Bank.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324).

Athey, S., & Imbens, G. (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects. arXiv preprint arXiv:1504.01132.

Anselin, Luc. *Spatial econometrics: methods and models.* Vol. 4. Springer Science & Business Media, 2013.

Anselin, Luc, et al. "Simple diagnostic tests for spatial dependence." *Regional science and urban economics* 26.1 (1996): 77-104.

H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded Up Robust Features. Lecture Notes in Computer Science, 3951:404–417, 2006

Belloni, Alexandre and Chernozhukov, V. (2013). "Least squares after model selection in high-dimensional sparse models" *Bernoulli*. 19(2).

Besley, T., & Ghatak, M. (2006). "Public goods and economic development". *Understanding Poverty*. (pp. 285-302). Oxford: Oxford University Press.

N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition (CVPR), San Diego, CA, 2005, pp. 886-893.

Department of Census and Statistics and World Bank, 2015 "The Spatial Distribution of Poverty in Sri Lanka", available at: http://www.statistics.gov.lk/poverty/SpatialDistributionOfPoverty2012_13.pdf

Donaldson D., and Storeygard A. "Big Grids: Applications of Remote Sensing in Economics", *forthcoming, JEP*.

Drukker, David M., Ingmar R. Prucha, and Rafal Raciborski. "Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances." University of Maryland, Department of Economics (2011).

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro–level estimation of poverty and inequality. Econometrica, 71(1), 355-364.

Elbers, Chris, Peter F. Lanjouw, and Phillippe G. Leite. "Brazil within Brazil: Testing the poverty map methodology in Minas Gerais." World Bank Policy Research Working Paper Series, Vol (2008).

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. Photogrammetric Engineering and Remote Sensing, 63(6), 727-734.

Engstrom, R., Ashcroft, E., Jewell, H., & Rain, D. (2011, April). Using remotely sensed data to map variability in health and wealth indicators in Accra, Ghana. In Urban Remote Sensing Event (JURSE), 2011 Joint (pp. 145-148). IEEE.

Engstrom, R., Sandborn, A., Yu, Q. Burgdorfer, J., Stow, D., Weeks, J., and Graesser, J. (2015) Mapping Slums Using Spatial Features in Accra, Ghana. Joint Urban and Remote Sensing Event Proceedings (JURSE), Lausanne, Switzerland, 10.1109/JURSE.2015.7120494

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348-1360.

Foster, James; Joel Greer; Erik Thorbecke (1984). "A class of decomposable poverty measures". *Econometrica*. 3. 52: 761–766.

S. W. Smith, The scientist and engineer's guide to digital signal processing. San Diego, CA: California Technical Publishing, 1997.

Gentle, J. E., Härdle, W. K., & Mori, Y. (Eds.). (2012). Handbook of computational statistics: concepts and methods. Springer Science & Business Media.

J. Graesser, A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E. Bright, "Image based characterization of formal and informal neighborhoods in an urban landscape," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no.4, pp. 1164-1176, Jul, 2012.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. The American Economic Review, 102(2), 994-1028.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. The American Economic Review, 105(5), 491-495.

Gabor, D. (1946). Theory of Communication. Journal of the Optical Society of America-A, 2 (2), 1455-1471.

Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2015). Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life (No. w21778). National Bureau of Economic Research.

Huettner, Frank, and Marco Sunder. "Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values." Electronic Journal of Statistics 6 (2012): 1239-1250.

Israeli, Osnat. "A Shapley-based decomposition of the R-square of a linear regression." The Journal of Economic Inequality 5.2 (2007): 199-212.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. The American economic review, 102(4), 1508.

Mooney, D. F., Larson, J. A., Roberts, R. K., & English, B. C. (2009). Economics of the variable rate technology investment decision for agricultural sprayers. In Southern agricultural economics association annual meeting, Atlanta, Georgia, January.

Mullahy, J. (1998). Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of health economics*, 17(3), 247-281.

Mullainathan, S. (2014, August). Bugbears or legitimate threats?:(social) scientists' criticisms of machine learning?. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 4-4). ACM.

Newhouse, David; Suarez Becerra, Pablo; Doan, Dung. 2016. Sri Lanka Poverty and Welfare: Recent Progress and Remaining Challenges. World Bank, Washington, DC. © World Bank. https://www.openknowledge.worldbank.org/handle/10986/23794 License: CC BY 3.0 IGO.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. Review of Economics and Statistics, 94(1), 20-36.

M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 1, no. 3, pp. 180-192, Oct, 2008.

Pinkovskiy, Maxim, and Xavier Sala-i-Martin. "Lights, Camera… Income! Illuminating the National Accounts-Household Surveys Debate." The Quarterly Journal of Economics 131.2 (2016): 579-631.

Sandborn, A. and Engstrom, R (In Press) Determining the Relationship Between Census Data and Spatial Features Derived From High Resolution Imagery in Accra, Ghana. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS) Special Issue on Urban Remote Sensing

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A. (2015). Data deprivation: another deprivation to end. World Bank Policy Research Working Paper, (7252).

Shorrocks, Anthony F. "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value." Journal of Economic Inequality (2013): 1-28.

Tucker CJ (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8: 127-150.

Watmough, G. R., Atkinson, P. M., Saikia, A., & Hutton, C. W. (2016). Understanding the Evidence Base for Poverty–Environment Relationships using Remotely Sensed Satellite Data: An Example from Assam, India. World Development, 78, 188-203.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790-794.

Marx, B., Stoker, T. M., & Suri, T. (2013). The Political Economy of Ethnicity and Property Rights in Slums: Evidence from Kenya.

Varian, H. R. (2014). Big data: New tricks for econometrics. The Journal of Economic Perspectives, 28(2), 3-27.

W. P. Yu, G. W. Chu, and M. J. Chung, "A robust line extraction method by unsupervised line clustering," Pattern Recognition, vol. 32, no. 4, pp. 529-546, Apr, 1999.

L. Wang, and D. He, “Texture classification using texture spectrum,” Pattern Recognition, vol. 23, no. 8, pp. 905-910, 1990.

Wong, T. H., Mansor, S. B., Mispan, M. R., Ahmad, N., & Sulaiman, W. N. A. (2003, May). Feature extraction based on object oriented analysis. In*Proceedings of ATC 2003 Conference* (Vol. 2021).

Tables

Table 1: Village (Grama Niladhari) Summary Statistics

	Mean	Sd	Min	Max
Economic Well-Being				
Avg Consumption in Rs	10274.2	3052.7	4881.9	21077
Avg Log Consumption	9.19	0.28	8.49	9.96
Rel. Pov. Rate at 10% Nat. Cons.	0.0903	0.066	0.0023	0.39
Rel. Pov. Rate at 40% Nat. Cons.	0.332	0.16	0.035	0.8
Geographic Descriptors				
log Area (square meters)	14.73	1.01	12.1	18
= 1 if urban	0.304	0.46	0	1
province==[1] Western	0.587	0.49	0	1
province==[3] Southern	0.255	0.44	0	1
province==[6] North-Western	0.0643	0.25	0	1
province==[7] North-Central	0.0155	0.12	0	1
province==[8] UVA	0.0782	0.27	0	1
Agricultural Land				
% of GN area that is agriculture	16.8	0.15	0	94
% of GN agriculture that is paddy	44.4	37.5	0	100
% of GN agriculture that is plantation	46.38	37.8	0	100
% of Total GN area that is paddy	8.629	10.9	0	74.7
% of Total GN area that is plantation	8.168	11	0	94.1
Cars				
log number of cars	3.123	1.44	0	8.3
Total cars divided by total road length	0.00556	0.01	0	0.17
Total cars divided by total GN Area	3.77E-05	0.00007	0	0.00093
Building Density and Vegetation				
% of area with buildings	7.817	6.82	0.13	33.9
% shadows (building height) covering valid area	6.509	6.01	0.31	34.9
Vegetation Index (NDVI), mean, scale 64	0.427	0.21	0	0.86
Vegetation Index (NDVI), mean, scale 8	0.566	0.24	0	0.99
Shadows				
ln shadow pixels (building height)	12.96	1.04	7.31	17.6
ln Number of Buildings	6.90	0.92	0	9.3
Road variables				
log of Sum of length of roads	9.445	0.94	1.47	13.1
fraction of roads paved	38.3	28.7	0	100
ln length airport roads	0.013	0.33	0	9.25
ln length railroads	1.098	2.67	0	10.8
Roof type				
Fraction of total roofs that are clay	36.5	22	0	100
Fraction of total roofs that are aluminum	14.08	7.06	0	71.9
Fraction of total roofs are asbestos	7.766	11.3	0	71.2
Textural and spectral characteristics				

Pantex (human settlements), mean	0.627	0.54	0.02	2.94
Histogram of Oriented Gradients (scale 64m), mean	3509.4	2070.3	129.1	10381
Linear Binary Pattern Moments (scale 32m), mean	49.5	1.1	18.1	49.5
Line support regions (scale 8m), mean	0.00836	0.004	-2E-07	0.035
Gabor filter (scale 64m), mean	0.469	0.28	0.014	1.3
Fourier transform, mean	84.34	17.8	4.51	113.4
SURF (scale 16m), mean	12.06	7.77	0.13	31.6
Observations	1291			

Table 2: Prediction of Local Area Poverty Rates Using High-Res Spatial Features

	10% Poverty Rate		40% Poverty Rate		Average GN Log Per Capita Consumption	
	Coef	t	coef	t	coef	T
log Area (square meters)	0.020*	[2.52]	0.0093	[0.60]	-0.0079	[-0.31]
= 1 if urban	-0.023	[-1.80]	-0.037	[-1.06]	0.08	[1.18]
% of GN area that is agriculture	-0.00025	[-1.04]	-0.00017	[-0.27]		
% of GN agriculture that is paddy	-0.00033**	[-2.97]	-0.00087**	[-2.97]	0.0014**	[2.92]
% of GN agriculture that is plantation	-0.00021**	[-2.84]	-0.00059*	[-2.66]	0.0012**	[2.72]
% of Total GN area that is paddy	-0.00019	[-0.58]	-0.00083	[-1.10]	0.0016*	[2.10]
Total cars divided by total road length	-0.31	[-1.17]				
Total cars divided by total GN Area	29.6	[0.54]				
log number of cars	-0.0059	[-0.89]	-0.015	[-1.39]	0.024	[1.60]
log sum of length of roads	-0.020***	[-3.64]	-0.027*	[-2.32]	0.033	[1.67]
fraction of roads paved	-0.00035***	[-4.24]	-0.00079**	[-3.24]	0.0014**	[3.06]
ln length airport roads	-0.0051	[-1.45]			0.022	[1.52]
ln length railroads	0.00098	[1.31]			-0.0046	[-1.26]
% of area with buildings	-0.0027*	[-2.31]	-0.0093*	[-2.34]	0.020*	[2.56]
log of Total count of buildings in GN	-0.0090**	[-2.71]	-0.019*	[-2.05]	0.029	[1.70]
Vegetation Index (NDVI), mean, scale 64	0.061*	[2.20]	0.14**	[2.94]	-0.21**	[-2.93]
Vegetation Index (NDVI), mean, scale 8	-0.064**	[-2.80]				
% shadows (building height)	0.0022*	[2.04]	0.0064*	[2.18]	-0.013*	[-2.27]
ln shadow pixels (building height)	0.016*	[2.51]	0.039*	[2.64]	-0.047	[-1.95]
Fraction of total roofs that are clay	0.00077**	[3.35]	0.0017**	[3.25]	-0.0027**	[-3.15]
Fraction of total roofs that are aluminum	0.00091***	[3.63]	0.0022**	[3.15]	-0.0040**	[-3.15]
Fraction of total roofs are asbestos	-0.00033	[-1.08]				
Linear Binary Pattern Moments (scale 32m)	0.0021**	[2.91]	0.0090***	[5.53]	-0.017***	[-5.92]
Line support regions (scale 8m), mean	-0.66	[-0.87]				
Gabor filter (scale 64m) mean	-0.052	[-1.53]				
Fourier transform, mean	0.0017**	[3.42]				
SURF (scale 16m), mean	-0.0014	[-0.94]	-0.001	[-0.59]	0.0034	[1.06]
Constant	-0.32**	[-3.03]	-0.31	[-1.43]	10.1***	[29.9]
Observations	1291		1291		1291	
R-sq	0.610		0.618		0.608	
R-sq Adj.	0.602		0.613		0.602	
Out-of-Sample R-sq	0.588		0.605		0.594	
Mean Absolute Error	0.032		0.078		0.139	

Notes: Unit of observation is Grama Niladhari (GN) division. Variables were selected using Lasso regularization from the candidate set of variables shown in table 1.
* p<0.05, ** p<0.01, *** p<0.001

Table 3: Marginal Effects of One Standard Deviation Change

	10% poverty rate	40% poverty rate	Average log consumption per capita	Variables
Area	2.1 pp *	0.9 pp	-0.008	Area
Urban	-1.0 pp	-1.7 pp	0.037	Urban Dummy
Agricultural land	-0.4 pp	-0.3 pp		% of GN area that is agriculture
Paddy land	-0.6 pp *	-1.9 pp **	0.026 *	All three variables combined
				% agriculture that is paddy
				% agriculture that is plantation (-)
				% GN area that is paddy
Cars	-1.2 pp			All three variables combined
				Total cars divided by total road length
				Total cars divided by total GN Area
		-2.2 pp	0.035	log number of cars
Road variables				
	-1.9 pp ***	-2.5 pp *	0.031	log sum of length
	-1.0 pp ***	-2.3 pp **	0.040 **	Fraction paved
	-0.2 pp		0.007	Log length of airport runway
	0.3 pp		-0.012	Log sum of railroads
Building Density	-2.7 pp **	-8.1 pp **	0.162 **	% of area with buildings and log of total count of buildings in GN combined
Vegetation	-0.2 pp			Both vegetation (NDVI) variables combined
		2.9 pp **	-0.044 **	Vegetation Index (NDVI), mean, scale 64
				Vegetation Index (NDVI), mean, scale 8
Shadows	3.0 pp ***	7.9 pp ***	-0.128 ***	% shadows (building height) and ln shadow pixels combined
Roofs				
	1.7 pp **	3.8 pp **	-0.06 **	Fraction of total roofs that are clay
	0.6 pp ***	1.6 pp **	-0.028 **	Fraction of total roofs that are aluminum
	-0.4 pp			Fraction of total roofs that are asbestos
Spatial Features				
	0.2 pp **	1.0 pp ***	-0.019 ***	Linear Binary Pattern Moments
	-0.3 pp			Line support regions
	3.1 pp			Fourier transform
	-1.5 pp			Gabor filter
	-1.1 pp **	-0.8 pp	0.026	SURF

Notes: Tables gives estimated marginal effect of a one standard deviation change in variable or variables listed in right column. For example, the combined marginal effect of a one standard deviation in all three cars variables on the 10 percent relative poverty rate is a reduction of 1.2 percentage points. Variables excluded from 40 percent poverty and log consumption models, as shown in Table 2, are also excluded when calculating marginal effects for those dependent variables. For agricultural land, % of GN that is plantation is subtracted from the sum of % GN agriculture that is paddy and % total GN area that is paddy. * p<0.05, ** p<0.01, *** p<0.001

Table 4: Shapley Decomposition of Share of Variance Explained (R^2) by High Resolution Spatial Feature Subgroup

	10% Poverty Rate	40% Poverty Rate	Average log predicted per capita consumption in GN
Area	10.4	8.3	8.4
Urban	9.4	9.7	10.8
Agricultural land	0.9	1.0	
Paddy land	3.8	4.6	4.1
Cars	7.3	5.6	4.6
Building density	14.8	19.5	22.5
Vegetation	8.0	6.2	4.4
Shadows	14.4	14.1	14.0
Road variables	9.4	7.7	9.8
Roof Type	10.4	8.3	8.4
Texture variables	9.4	9.7	10.8
Observations	1291	1291	1291
R-sq	0.610	0.618	0.608

Notes: Agricultural variables include fraction agriculture plantation, fraction agriculture paddy, and fraction of GN area that is plantation. Car variables include log of car count, and cars per total road length. Building density variables include log of developed area, shadow count (building height proxy), fraction of GN developed, fraction covered by shadow, NDVI at scales 64 and 8. Road variables include log of unpaved road length, log of paved roads narrower than 5m, log of paved roads 5m+, log of airport roads, log of railroad length, and fraction of roads paved. Roof variables include count of roofs by type: clay, aluminum, asbestos, grey cement, and fraction of roofs of same type. Texture variables include Fourier series, Gabor, histogram of oriented gradients, Local Binary Pattern Moments mean and standard deviation, line support regions, and SURF.

Table 5: Linear Model Estimates Night Lights on Small Area Poverty/Average GN Consumption

	(1) **10% Pov. Rate**	(2) **40% Pov. Rate**	(3) **Avg. Income**	(4) **10% Pov. Rate**	(5) **40% Pov. Rate**	(6) **Avg. per capita consumption**
Night Lights 2012	-0.583***	-1.546**	2.922**	-0.0383	-0.0898	0.186
	(-3.53)	(-3.38)	(3.32)	(-0.79)	(-0.67)	(0.64)
Observations	1291	1291	1291	1291	1291	1291
R-sq	0.109	0.131	0.147	0.000868	0.000842	0.00103
R-sq Adj.	0.108	0.130	0.146	0.0000932	0.0000671	0.000258
R-sq within				0.000868	0.000842	0.00103
R-sq between				0.372	0.448	0.527
R-sq overall				0.109	0.131	0.147
Divisional Secretariat FEs	No	No	No	Yes	Yes	Yes

Unit of observation is Grama Niladhari (GN) Division.
All models include a regression constant which is omitted from the table.
* p < 0.05, ** p < 0.01, *** p < 0.001

Table 6: Shapley Decomposition of Share of Variance Explained (R^2) by High Resolution Spatial Feature Subgroup in Addition to Night Time Lights

	10% Poverty Rate	40% Poverty Rate	Average log predicted per capita consumption in GN
Area	10.2	8.1	8.0
Urban	8.7	8.7	9.5
Agricultural land	0.9	1.0	3.3
Paddy land	3.3	3.8	
Cars	6.7	5.1	4.0
Buildings	13.0	16.7	19.0
Vegetation	8.0	6.0	4.1
Shadows	12.1	13.0	10.6
Road variables	8.0	8.0	8.5
Roof Type	13.0	12.0	11.7
Texture variables	8.5	7.1	8.9
Night Time Lights variables	7.6	10.6	12.1
Observations	1291	1291	1291
R-sq	0.621	0.636	0.632

Notes: Night time lights category includes the following transformations of night time lights: average, squared, cubed, and standard deviation. Agricultural variables include fraction agriculture plantation, fraction agriculture paddy, and fraction of GN area that is plantation. Car variables include log of car count, and cars per total road length. Building density variables include log of developed area, shadow count (building height proxy), fraction of GN developed, fraction covered by shadow, NDVI at scales 64 and 8. Road variables include log of unpaved road length, log of paved roads narrower than 5m, log of paved roads 5m+, log of airport roads, log of railroad length, and fraction of roads paved. Roof variables include count of roofs by type: clay, aluminum, asbestos, grey cement, and fraction of roofs of same type. Texture variables include Fourier series, Gabor, histogram of oriented gradients, Local Binary Pattern Moments mean and standard deviation, line support regions, and SURF.

Table 7: Marginal Effects of One Standard Deviation Change for Urban and Rural Models

	Urban	Rural	Variables
Area		-0.032	Area
Agricultural land		0.018 *	% of GN area that is agriculture
Paddy land	0.045		Combined Paddy and plantation
		0.026 **	% of GN agriculture that is paddy
			% of GN agriculture that is plantation (-)
Cars			
	0.093 ***	0.029 ***	Log car count
Road variables			
		0.030 *	log sum of length
	0.041	0.029 ***	Fraction paved
		0.011 ***	Log length of airport runway
	-0.02		Log sum of railroads
Building Density			
	0.186 ***		Both building density variables
			% of area with buildings,
		0.038 ***	log of Total count of buildings in GN
Vegetation	0.041	-0.060 ***	Vegetation Index (NDVI), mean, scale 64
Shadows			
	-0.107 **		% shadows
		-0.061 ***	ln shadow pixels
Roofs			
	0.036	-0.084 ***	Fraction of total roofs that are clay
	-0.022	-0.037 ***	Fraction of total roofs that are aluminum
		-0.021 *	Fraction of total roofs that are asbestos
Spatial Features			
		-0.018	Linear Binary Pattern Moments
	-0.006		Line support regions
	-0.058		Fourier transform
		0.075 ***	Pantex
			SURF
Observations	393	898	
R-sq	0.446	0.656	
R-sq Adj.	0.427	0.650	
Out-of-Sample R-sq	0.412	0.641	
Mean Absolute Error	0.145	0.113	

Notes: Tables gives estimated marginal effect of a one standard deviation change in variable or variables listed in right column. For example, the combined marginal effect of a one standard deviation in all three cars variables on the 10 percent relative poverty rate is a reduction of 1.2 percentage points. Variables excluded from 40 percent poverty and log consumption models, as shown in Table 2, are also excluded when calculating marginal effects for those dependent variables. For agricultural land, % of GN that is plantation is subtracted from the sum of % GN agriculture that is paddy and % total GN area that is paddy. * p<0.05, ** p<0.01, *** p<0.001

Table 8: Model performance for prediction of average log per capita consumption at different points in the welfare distribution consumption

	Bottom 20%	**Bottom 40%**	**Bottom 60%**	**Bottom 80%**	**Full Sample**
Observations	259	517	775	1033	1291
R-sq	0.551	0.454	0.474	0.509	0.608
Adjusted R-sq	0.52	0.436	0.461	0.5	0.602
Out of sample	0.487	0.425	0.447	0.475	0.595
Mean Absolute Error	0.064	0.0774	0.0909	0.115	0.139
Mean log per capita income	8.83	8.95	9.00	9.09	9.16
Standard deviation	0.11	0.13	0.15	0.20	0.28

Notes: Table reports model performance statistics for the national model for different subsamples of the bottom portion of the GN Division welfare distribution. The dependent variable is average predicted log GN per capita consumption. The set of independent variables is given in Table 2 and is identical for all subsamples. Subsamples are constructed using the unweighted quintiles of GN average predicted per capita consumption. The rightmost column is identical to the results reported in the right column of Table 2.

Table 9: MLE Estimation Correcting for Spatial Autoregression

	Average log Village Consumption (Income)	
	coef	t
log Area (square meters)	-0.046***	[-4.01]
= 1 if urban	0.048+	[1.96]
% of GN area that is agriculture	0.00022	[0.42]
% of GN agriculture that is paddy	0.00046+	[1.74]
% of GN agriculture that is plantation	0.00076**	[3.09]
% of Total GN area that is paddy	0.00057	[0.79]
Total cars divided by total road length	-0.93	[-1.20]
Total cars divided by total GN Area	401.4*	[2.28]
log number of cars	0.020***	[3.57]
% of area with buildings	0.0083***	[4.19]
log of Total count of buildings in GN	0.012	[1.23]
Vegetation Index (NDVI), mean, scale 64	0.071	[1.54]
Vegetation Index (NDVI), mean, scale 8	-0.042	[-0.67]
log of Sum of length of roads	0.029**	[2.70]
fraction of roads paved	0.0012***	[6.00]
ln length airport roads	0.0052	[1.50]
ln length railroads	-0.00092	[-0.48]
Fraction of total roofs that are clay	-0.0025***	[-5.83]
Fraction of total roofs that are aluminum	-0.0034***	[-4.92]
Fraction of total roofs are asbestos	0.0014*	[2.26]
Linear Binary Pattern Moments (scale 32m), mean	-0.0080***	[-3.38]
Line support regions (scale 8m), mean	-1.25	[-0.71]
Gabor filter (scale 64m) mean	-0.053	[-0.92]
Fourier transform, mean	-0.0030***	[-3.61]
SURF (scale 16m), mean	0.0052*	[2.24]
Constant	9.74***	[51.6]
Observations	1287	

Notes: Standard errors have been corrected according to Conley (1999, 2008), with model estimation via GMM. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Table 10: Estimating Poverty Gap Using High Res Features

	Poverty Gap (FGT1 - 10%)		Poverty Gap (FGT1 - 40%)	
	coef	t	coef	t
log Area (square km)	0.0060**	[2.84]	0.0063	[1.02]
= 1 if urban	-0.0063	[-2.00]	-0.013	[-1.05]
% of GN area that is agriculture	-0.000081	[-1.29]	-0.00018	[-0.76]
% of GN agriculture that is paddy	-0.000087**	[-3.24]	-0.00033**	[-3.10]
% of GN agriculture that is plantation	-0.000053**	[-2.91]	-0.00021*	[-2.63]
% of Total GN area that is paddy	-2.3E-05	[-0.29]	-0.00025	[-0.88]
Total cars divided by total road length	-0.09	[-1.32]		
Total cars divided by total GN Area	9.55	[0.72]		
log number of cars	-0.0014	[-0.83]	-0.0058	[-1.24]
log of Sum of length of roads	-0.0049**	[-2.97]	-0.011*	[-2.48]
fraction of roads paved	-0.000077**	[-3.37]	-0.00023*	[-2.67]
ln length airport roads	-0.00027	[-0.89]		
ln length railroads	0.00026	[1.35]		
% of area with buildings	-0.00062*	[-2.16]	-0.0028*	[-2.04]
% shadows (building height) covering valid area	0.00053	[1.76]	0.0017	[1.54]
ln shadow pixels (building height)	0.0037*	[2.19]	0.016*	[2.68]
Fraction of total roofs that are clay	0.00020**	[2.96]	0.00070**	[3.12]
Fraction of total roofs that are aluminum	0.00024**	[3.31]	0.00084**	[3.19]
Fraction of total roofs are asbestos	-9.1E-05	[-1.14]		
log of Total count of buildings in GN	-0.0022*	[-2.62]	-0.0073*	[-2.09]
Vegetation Index (NDVI), mean, scale 64	0.017*	[2.33]	0.056**	[2.88]
Vegetation Index (NDVI), mean, scale 8	-0.019**	[-2.95]		
Linear Binary Pattern Moments (scale 32m)	0.00048*	[2.55]	0.0029***	[4.87]
Line support regions (scale 8m), mean	-0.27	[-1.39]		
Gabor filter (scale 64m) mean	-0.016	[-1.78]		
Fourier transform, mean	0.00046**	[3.44]		
SURF (scale 16m), mean	-0.00025	[-0.67]	-0.0001	[-0.15]
Constant	-0.093**	[-3.41]	-0.17+	[-2.00]
Observations	1234		1234	
R-sq	0.5884		0.6097	
R-sq Adj.	0.5792		0.6039	

* p<0.05, ** p<0.01, *** p<0.001

Table 11: Model Performance Using Simulated Reduced Census Sampling

Sample of villages (GNs) for model training	Sample of households within GNs	In-Sample R^2	Out of Sample R^2	Normalized Mean Absolute Error (NMAE)
Using 10% National Poverty Line As Dependent Variable				
100%	100%	0.6325	-	0.3272
	50%	0.6316	-	0.3271
	25%	0.6287	-	0.3283
50%	100%	0.6058	0.6379	0.3355
	50%	0.6061	0.6378	0.3348
	25%	0.6092	0.6441	0.3342
25%	100%	0.5995	0.6139	0.3393
	50%	0.5990	0.6167	0.3395
	25%	0.5948	0.6212	0.3392
Using 40% National Poverty Line As Dependent Variable				
100%	100%	0.6215	-	0.2300
	50%	0.6213	-	0.2300
	25%	0.6209	-	0.2303
50%	100%	0.6097	0.6104	0.2348
	50%	0.6097	0.6118	0.2346
	25%	0.6098	0.6154	0.2340
25%	100%	0.5947	0.6132	0.2363
	50%	0.5939	0.6152	0.2357
	25%	0.5923	0.6178	0.2346
Using Average Per Capita Consumption as Dependent Variable				
100%	100%	0.6081	-	0.014778
	50%	0.6080	-	0.014783
	25%	0.6077	-	0.014782
50%	100%	0.5989	0.5964	0.01499
	50%	0.5987	0.5982	0.01498
	25%	0.5984	0.6014	0.01493
25%	100%	0.5880	0.5943	0.01524
	50%	0.5874	0.5954	0.01520
	25%	0.5866	0.5966	0.01515

Notes: This table simulates estimation error when using a reduced Census size. "Sample of villages" refers to the percentage of the villages within the Census used to train the model. "Sample of households within GN" refers to the number of households within the sampled GNs used to calculate the income or poverty rate statistic. In-sample (out-of-sample) R^2 reports the coefficient of determination for the data used in the training (test) sample. Normalized mean absolute error (NMAE) reports the mean average error rate divided by the average income/poverty rate, such that the statistic gives the average absolute error expressed as a percentage of the income/poverty rate.

Table 12: Estimated Poverty Extrapolation Performance, Using DS Leave-One-Out Cross-Validation (LOOCV)

	Average predicted per capita consumption in GN	10% Poverty Rate	40% Poverty Rate
Normalized Root Mean Squared Error (NRMSE)	0.084	0.560	0.363
Normalized Mean Absolute Error (NMAE)	0.024	0.405	0.277
Spearman Rank Correlation Between Predicted and True Poverty Rates	0.698	0.694	0.679

Notes: Table presents out of sample estimates for of extrapolated poverty rate prediction into withheld Divisional Secretariat (DS) administrative districts. We estimate 47 models, each time withholding one of 47 DS units to reserve as an out of sample test. Using the relationship between poverty and satellite variables estimated using the training data, we predict into the withheld DS – so called "leave-one-out" cross-validation (LOOCV).

Appendix A: Description of Imagery and Extraction of Object and Texture Features from High Resolution Satellite Features

6.1 Details on Satellite Imagery

The satellite imagery consists of 55 unique "scenes" purchased from Digital Globe, covering areas specified in our sample area.[31] Each "scene" is an individual image captured by a particular sensor at a particular time. Images were acquired by three different sensors: Worldview 2, GeoEye 1, and Quickbird 2. These sensors have a spatial resolution of $0.46m^2$, $0.41m^2$, and $0.61m^2$, respectively in the panchromatic band and $1.84m^2$, $1.65m^2$, $2.4m^2$ respectively in the multi-spectral bands. Pre-processing of imagery included pan-sharpening, ortho-rectification, and image mosaicking.

6.2 Details on Extraction of Object Based Features

Object features were classified using the assistance of two technical partners: Orbital Insight and LandInfo. Orbital Insight produced object classification for three variables: The share of the GN division that is built-up (i.e. consists of buildings), the number of cars in the GN, and the share of pixels in the GN that were identified as shadow pixels, which is a proxy for the gross floor area, or height, of buildings. The classification method used by Orbital Insight is similar to Krizhevsky, Sutskever, and Hinton (2012) which utilizes convolutional neural networks (CNN) to build object predictions from raw imagery. LandInfo classified the remaining objects, which included roof type, paved and unpaved roads of different widths, railroads, and the type of agriculture. Landinfo used a combination the Trimble eCognitition and Erdas Imagine software platforms to classify objects, except for roads, which were classified using visual interpretation.

The CNN classification algorithm used by Orbital Insight involved four steps:

1. Ingestion/Tiling
2. Model Development
3. Classifying All Pixels Using the Trained Model
4. Aggregating Prediction Results to GN Division level

The tiling stage split the large images into many small images or tiles, in order to make the modeling computationally scalable, as each tile could be distributed to a different GPU core for greater efficiency. In the model development stage, the classification model was trained and tuned. Model building began by manually classifying or labeling a sub-sample of the imagery as a positive or negative value for a given object using a crowdsourced campaign. The classified data was split into an 80% training and a 20% testing set, where the training set was used to build the model. This allowed sample prediction metrics, presented below, to be calculated using the withheld test set. Training was run for 60,000 iterations using the Nesterov solver method, a variant of stochastic gradient descent.

To get a sense of the accuracy, Figure A1 shows the receiver operator characteristics, or ROC curve, summarizing the classification accuracy of the developed area building classifier. The ROC curve

[31] Particular thanks to Digital Globe is due for their "Seeing the World" program, which offered very high spatial resolution imagery at reduced rates for non-commercial purposes.

presents the true positive rate on the y-axis, against the false positive rate on the x axis as the discriminant threshold is varied. A classifier that is no better than random would correspond to the 45 degree line, while a perfect classifier would correspond to the point (0,1) in the ROC space. Improvements in classifier accuracy are shown as the curve moves up and to the left. The ROC curve suggests the classification algorithm is highly accurate, corresponding to a 90% accuracy rate overall. Once the model was full trained, the team applied the trained model to the full set of imagery. The results were then summarized at the GN level.[32]

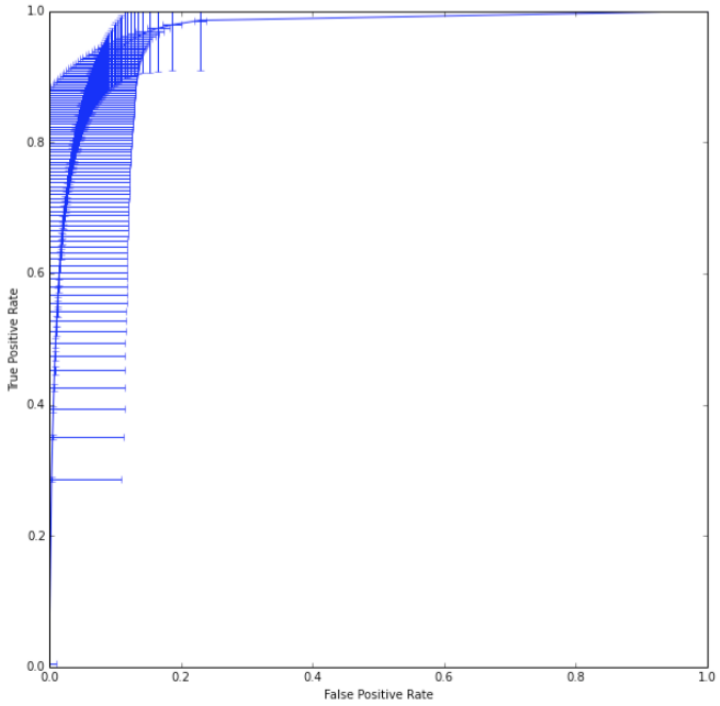

Figure A1 ROC Curve for Developed Area (Buildings) Classifier

LandInfo extracted roof type, paved and unpaved roads of different widths, railroads, and the type of agriculture, using a combination of an object-based image analysis (OBIA) methodology utilizing the Trimble eCognition software platform (Benz et al., 2004) and visual interpretation. OBIA is a common classification methodology applied to HSRI because it is designed to map objects that are larger than the pixel size (Blaschke, 2010). For this OBIA analysis, the first step was to segment the image into polygons that represent the features of interest based on spectral and spatial homogeneity of those features. The segments were created based on the degree of homogeneity using a range of scale, shape, and other parameters. Once an image was segmented, the objects themselves were then classified, using Erdas Imagine Software. Using this methodology, all of the roofs of individual buildings within the imagery were extracted and the type of roof was classified as asbestos, aluminum, clay, or grey roofs, in part based on validation of selected roofs in Colombo. The roof types of a few selected buildings were validated by direct observation. Roads and agriculture were extracted using a combination of eCognition and manual

[32] The GN shapefile was in turn based on the GN shapefile provided by DCS, modified to correct manually identified errors.

visual interpretation. Visual interpretation was required because the roads were covered by trees and in shadow, making automated detection difficult. In addition, the imagery was not always taken during the growing season, which along with the small, irregular shapes of the agricultural plots made automated extraction impractical.

Figure Figure A2 shows an example OBIA classification for roads, railroads, and road width. The road network in the village has been mapped in detail, showing not just major roads, but minor roads. Not shown are the road type, whether the road is paved or not. These features combine to indicate the extent to which a given area is accessible by road. Figure A3 shows an example roof type classification, where roof type is distinguished not just by material but by material shading.

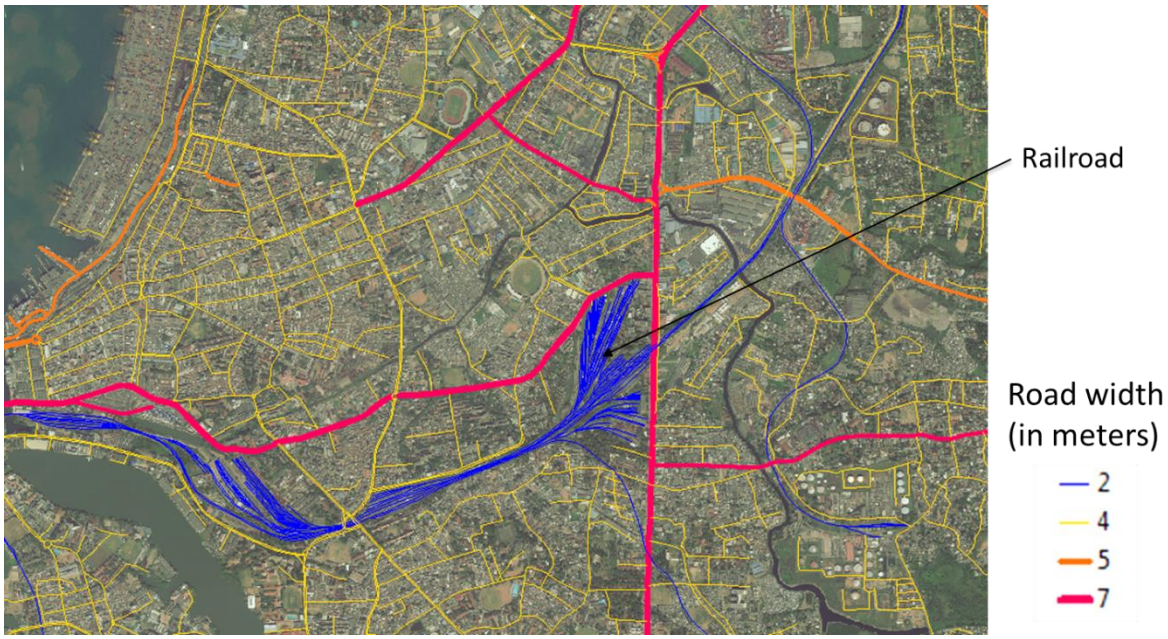

Figure A2: Example Roads and Railroads Classification

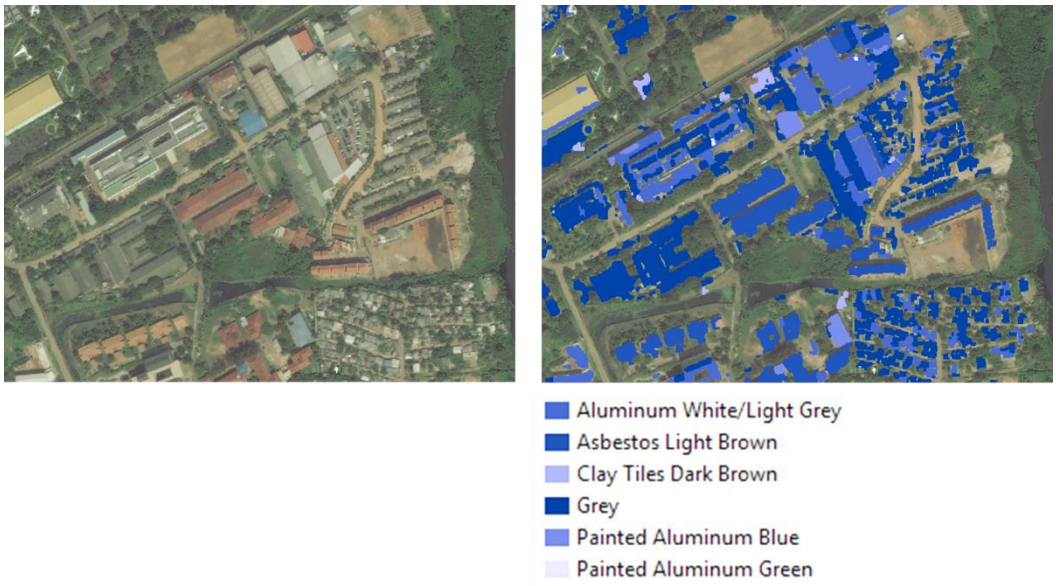

Figure A3: Example Roof Type Classification

6.3 Details on Extraction of Spectral and Textural features

Textural and spectral features were created using a block size of eight pixels and scales of 8, 16, 32, and 64 meters using a methodology similar to Graesser et al. (2012). This resulted in an output image comprised of 165 bands at a spatial resolution of 12.8 or 16m depending on native, multispectral resolution of each of the sensors. The seven textural features calculated for this study were:

1. Histogram of oriented gradients (HOG), which captures edge orientations and sorts them into a histogram (Dalal and Triggs 2005).
2. PanTex, which is a built-up presence index derived from the grey-level co-occurrence matrix (GLCM) (Pesaresi et al. 2008).
3. Line support regions (LSR), which characterize line attributes (Yu et al. 1999)
4. Local binary patterns moments (LBPM), which define contiguous regions of pixel groups and sorts them into a histogram (Wang and He, 1990).
5. Fourier transform (FT) which examines pattern frequency across an image (Smith 1997).
6. Gabor, a linear Gaussian filter used for edge detection (Gabor 1946)
7. Speeded Up Robust Features (SURF), an algorithm that extracts key points (i.e., edges and corners) from an image through pyramidal Gaussian based decomposition (Bay et al., 2006).
8. The Normalized Difference Vegetation Index (NDVI), the most widely used vegetation index that provides information about the presence and abundance of vegetation (Tucker 1979).

Additional spectral features calculated were simply the means of the four individual bands, Blue, Green, and Near Infrared. Once the spatial and spectral features were calculated, the mean, standard deviation, and sum were determined for each GN Division. Previous research has indicated that these features are correlated with census data that indicate poverty such as slum conditions, population density, solid waste collection, unimproved sanitation (Sandborn and Engstrom, 2016) and to map informal and slum areas within cities (Graesser et al. 2012, Engstrom et. al. 2015).

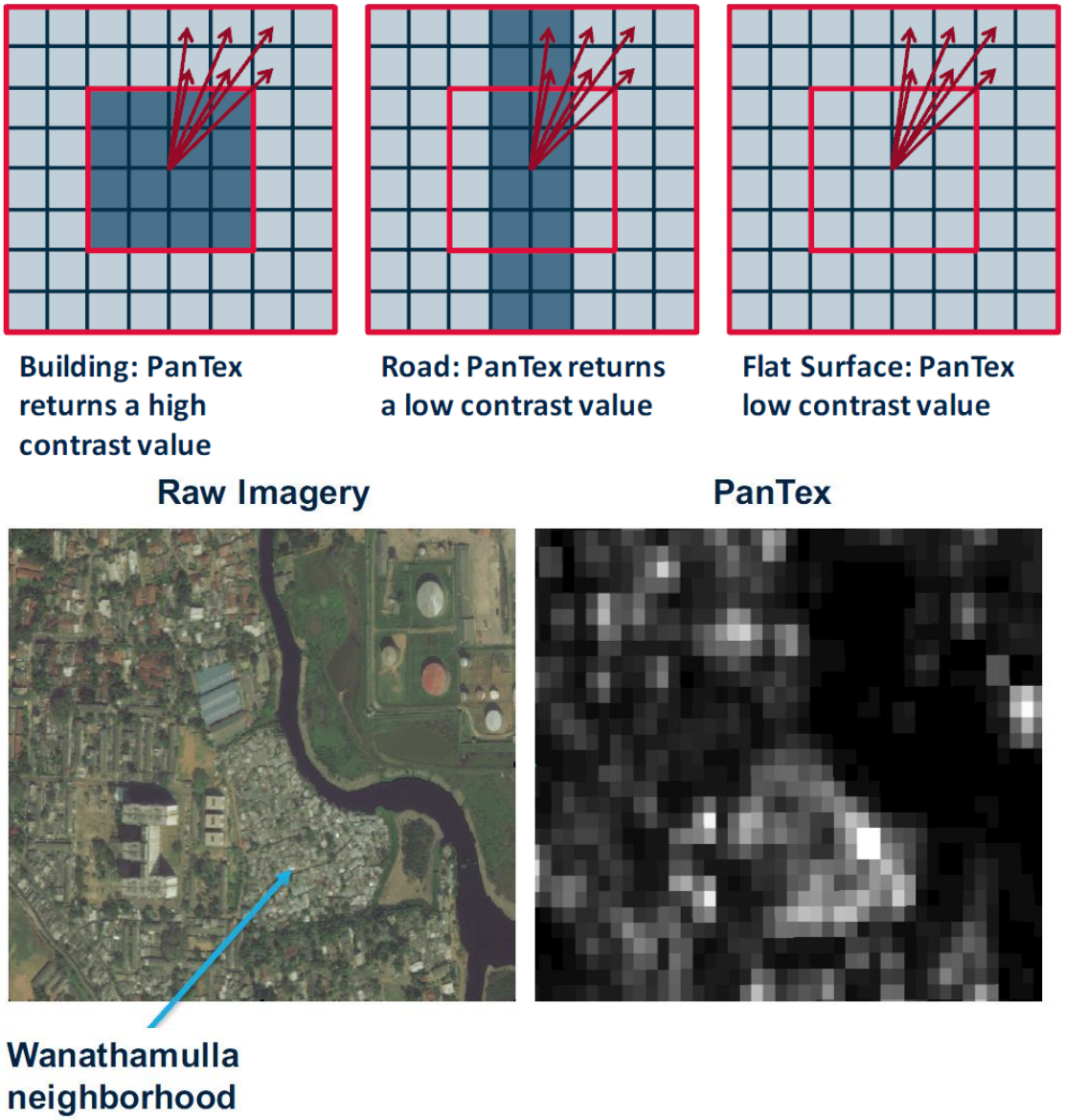

Figure A4: Pantex Classification Algorithm Description (top) and Example Classification (bottom right) Applied to Raw Imagery (bottom left)

Sampled divisional secretariats include Ambagamuwa, Ambalantota, Ambanpola, Bandaragama, Biyagama, Bulathsinhala, Colombo, Dehiwala, Devinuwara, Dodangoda, Doluwa, Dompe, Galle Four Gravets, Hali Ela, Hambantota, Homagama, Horana, Ingiriya, Kaduwela, Kalutara, Kamburupitiya, Katana, Kattankudy, Kelaniya, Kesbewa, Kirinda Puhulwella, Kolonnawa, Kotapola, Kothmale, Kurunegala, Madurawala, Maharagama, Malimbada, Manmunai North, Matara Four Gravets, Moratuwa, Nagoda, Negambo, Nuwara Eliya, Nuwaragam Palatha East, Padukka, Panadura, Panvila, Puttalam, Rathmalana, Rattota, Seethawaka, Sri Jayawardanapura Kotte, Thihagoda, Thimbirigasyaya, Tissamaharama, dapalatha, Udunuwara, Ukuwela, and Uva Paranagama

Appendix B: Information on Sri Lankan Administrative Divisions

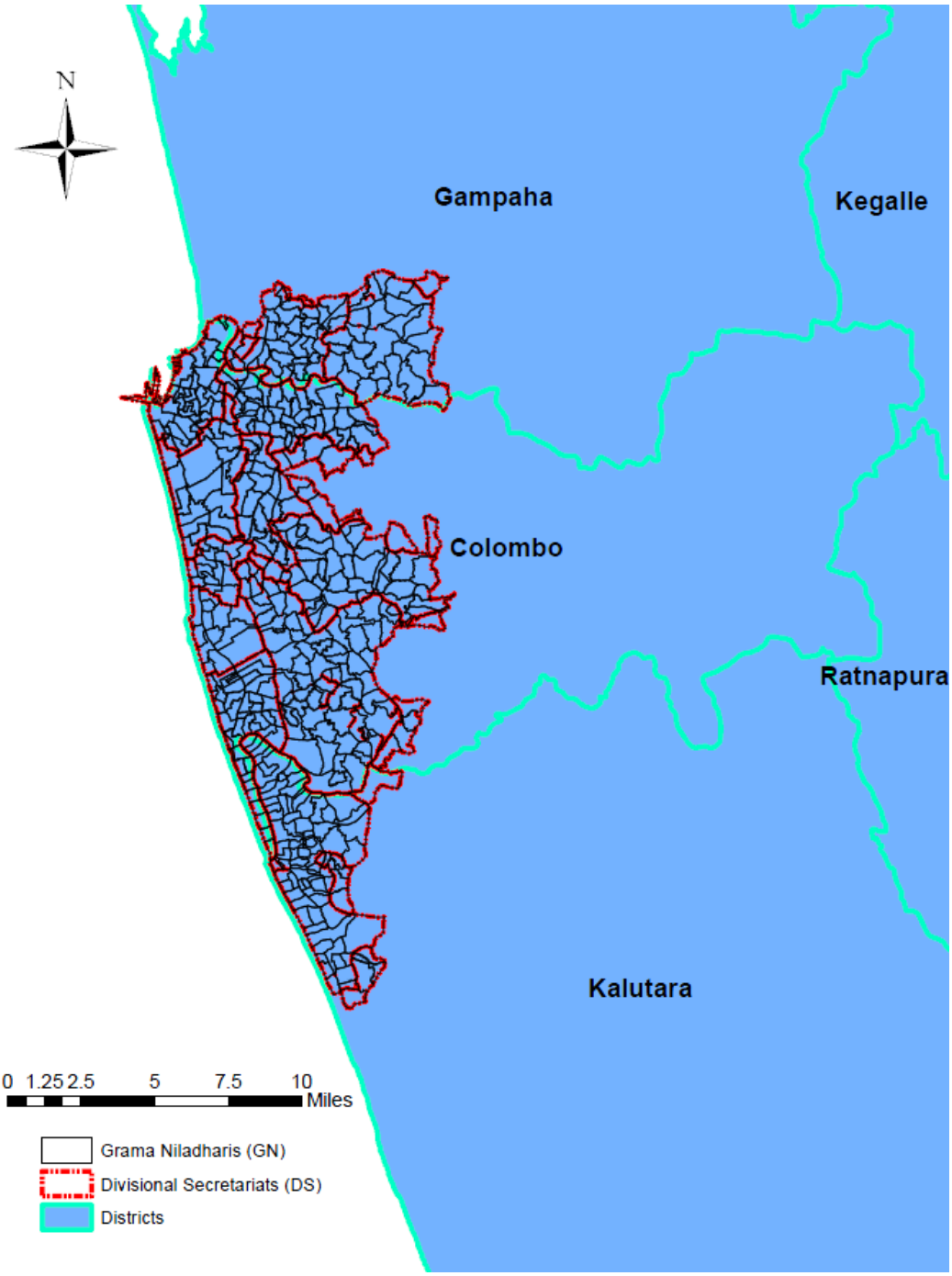

Figure B1: Graphical Depiction of Relative Size of Districts, Divisional Secretariats, and Grama Niladharis

Figure A2: Count of Administrative Divisions of Sri Lanka by Type

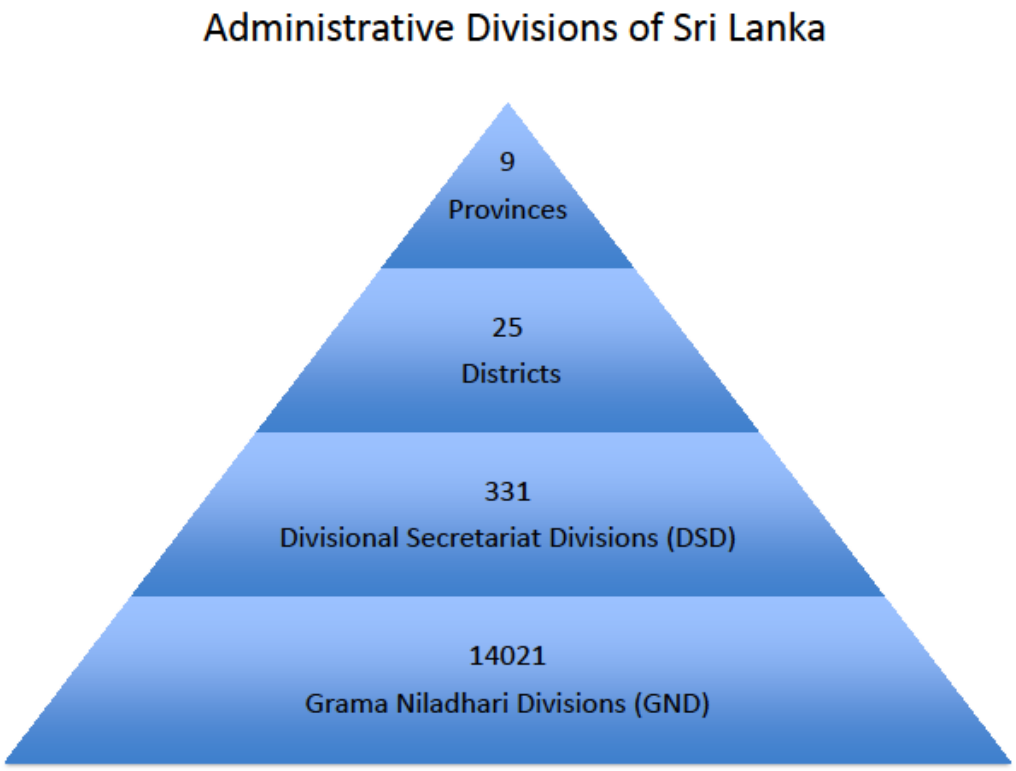

Figure B2: Count of Administrative Divisions of Sri Lanka by Type