

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

Ryan Engstrom, Jonathan Hersh, and David Newhouse

Abstract

Can features extracted from high spatial resolution satellite imagery accurately estimate poverty and economic well-being? The present study investigates this question by extracting both object and texture features from satellite images of Sri Lanka. These features are used to estimate poverty rates and average expected log consumption taken from small-area estimates derived from census data, for 1,291 administrative units. Features extracted include the number and density of buildings, the prevalence of building shadows (proxying building height), the number of cars, length of roads, type of agriculture, roof material, and several texture and spectral features. A linear regression model explains between 49 and 61 percent of the variation in average expected log consumption, and between 37 and 62 percent for poverty rates. Estimates remain accurate throughout the consumption distribution, and when extrapolating predictions into adjacent areas, although performance falls when using fewer households to calculate estimates of poverty and welfare.

JEL classification: I32, C50

Keywords: poverty estimation, satellite imagery, machine learning, big data, inequality

Ryan Engstrom is an associate professor of geography at George Washington University in Washington, DC; his email address is rengstro@gwu.edu. Jonathan Hersh (corresponding author) is an assistant professor of economics and management science at Chapman University in Orange, CA, and may be reached at hersh@chapman.edu; David Newhouse is a Senior Economist at the Poverty and Equity Global Practice at the World Bank. His email address is dnewhouse@worldbank.org.

This project benefited greatly from the comments of two anonymous referees and discussions with Sarah Antos, Ana Areias, Marianne Baxter, Sam Bazzi, Azer Bestavros, Jacob Bien, Kristen Butcher, John Byers, Pedro Conceição, Francisco Ferreira, Ray Fisman, Michael Gechter, Alex Guzey, Klaus-Peter Hellwig, Kristen Himelein, Selim Jahan, Matthew Kahn, Tariq Khokhar, Kala Krishna, Hannes Mueller, Trevor Monroe, Dilip Mookherjee, Vivian Peng, Pierre Perron, Hashem Pesararan, Bruno Sánchez-Andrade Nuño, Kiwako Sakamoto, Jacob Shapiro, David Shor, Benjamin Stewart, Andrew Whitby, Nat Wilcox, Nobuo Yoshida, and seminar participants at Boston University, Chapman University, University of Southern California, Penn State, Princeton University, UNDP, 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 the authors thank Dr. Amare Satharasinghe for authorizing the use of the Sri Lankan census data. Liang Xu and Cady Stringer 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.

1. Introduction

Despite the best efforts of national statistics offices and the international development community, local area estimates of poverty and economic welfare remain rare. 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 are typically available only once per decade at best. Furthermore, in many conflict areas safety concerns may prohibit survey data collection altogether.

Satellite imagery has generated considerable enthusiasm as a potential supplement to household data that can help fill these data gaps. In recent years, private companies such as DigitalGlobe (now Maxar) and Airbus have rapidly expanded the coverage and availability of High Spatial Resolution Imagery (HSRI), driving down commercial prices. The startup Planet currently operates hundreds of satellites with the goal of daily coverage of the entire planet at 3 to 5 meter spatial resolution per pixel. Continued technological advances are likely to further 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 to estimate and predict poverty rates at local levels. The study area of this paper covers 3,500 square kilometers in Sri Lanka, which contain 1,291 local administrative areas known as Grama Niladhari (GN) divisions. The study employs a two-step methodology in which, for each GN, it extracts meaningful object and texture features from the satellite images, and then uses these features to model poverty, average income, or an asset index of an area. Object features extracted include the number of cars, number and size of buildings, type of farmland (plantation or paddy), the type of roofs, the share of shadow pixels (building height proxy), road extent and road material, along with contextual measures. These features are identified using a combination of deep learning-based Convolutional Neural Networks (CNN) and classification of spectral and textural characteristics. 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.

The article investigates five main questions: 1) To what extent can variation in GN economic well-being—headcount poverty rates defined at the 10 and 40th percentiles of national income and average GN consumption—be explained by high spatial-resolution features? 2) Which features are most strongly correlated with these measures of well-being? 3) Do these features predict equally well in poor and rich GNs? 4) Can these models predict into areas different from those in which the model was estimated? Finally, 5) how robust is the prediction model to using a smaller number of households and a single simulation per household to generate ground-truth measures of GN poverty and welfare?

The study finds that 1) satellite features are highly predictive of economic well-being and explain between 35 and 60 percent of the variation in both GN average consumption and estimated poverty headcount rates; 2) 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; 3) accuracy declines only slightly in the poorest decile of villages (average consumption of \$4.67 per day); 4) predicting into geographically distinct areas sees a slight reduction in accuracy but remains relatively high; and v) the predictive power of the model is highly sensitive to the number of households and the number of simulations used to generate the “ground truth” training data.

This paper contributes to a growing literature exploring how remotely sensed data may be used to assess economic outcomes. The first paper, to the authors’ knowledge, that combined satellite and survey data for prediction used daytime imagery from Landsat to predict the area under corn and soybean cultivation in 12 Iowa counties (Battese, Harter, and Fuller 1988). Since then, 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 (Elvidge et al. 1997; Henderson, Storeygard, Weil 2012; Pinkovskiy and Sala-i-Martin 2016), although within a country NTL appears more strongly correlated with density than with 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 economic activity. 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 km².¹ Indeed, this study finds that NTL captures only 20 percent of the variation in poverty or income in the same area where high resolution spatial features capture 35–60 percent of the variation.

Daytime imagery has recently re-emerged as a practical source of information on welfare, in large part due to new developments in computer vision algorithms. Advances in Deep Learning such as Convolutional Neural Networks (CNN) have the capability to algorithmically classify 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. 2015b; Sandborn and Engstrom 2016; Engstrom et al. 2019). 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 this study refers to as contextual features, 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, such as images or text, to reveal information about human welfare (Donaldson and Storeygard 2016; Athey 2017; Gechter and Tsivanidis 2018). Glaeser et al. (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 perceived safety. They estimate a support-vector machine model and show that the fitted model can reliably predict block-level income in New York City. Jean et al. (2016) employ 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 explains an average of 46 percent of the variation in village per capita consumption, out of sample, across the four countries in which it was trained. Subsequent researchers estimated a direct end-to-end CNN to model poverty in Mexico (Babenko et al. 2017), Africa (Yeh et al. 2020) and Uganda (Ayush et al. 2020) without the NTL transfer learning stage.² While Jean et al.'s innovative use of daytime imagery substantially improves on the use of night-time lights alone, there are two problems with its applicability to poverty measurement. First, extensions of this approach in Haiti and Nepal (Head et al. 2017) show declines in predictive performance, suggesting the NTL step in the transfer learning process may be ill suited for many areas, especially those that are primarily dark when viewed through NTL. Second, the transfer learning method is not necessarily optimal for predicting very poor areas. 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 to about 0.12. In contrast, this study's method explains 48–55 percent of the

1 Pixel size can vary depending on the angle of the satellite relative to the ground site.

2 This current paper is also distinguished from a previous conference proceeding, which only uses spatial features and restricts the analysis to Colombo (Engstrom et al. 2017b).

variation in the poorest decile of villages.³ Head et al. (2017), comment on the transfer learning approach that it may be “possible that other approaches to feature engineering might be more successful than the brute force approach of the convolutional neural network.” This is precisely what this paper does.

Other researchers have used CNNs to predict an asset index as the outcome variable. Yeh et al 2020 train a CNN model directly to household survey data from 23 DHS surveys in Sub-Saharan Africa, using both daytime and night-time imagery. The prediction explains roughly 70 percent of survey-measured average wealth at the cluster level, out of sample. Furthermore, correlations between predictions at the district level and independent census data are only slightly weaker than the correlation between the survey data and the census data. Unlike this paper, however, Yeh et al 2020 only evaluates the ability of satellite imagery to predict a measure of asset wealth, whereas this paper predicts log per capita consumption and poverty rates directly. In addition, Yeh et al 2020 is limited to medium-resolution publicly available imagery from the Landsat satellite. The present paper also differs in that its primary results are based on regression models of distinct interpretable features, and in that way is more similar to Ayush et al (2020).

This paper differs in four significant ways from Jean et al. (2016), as well as related articles that predict differences in village wealth using satellite imagery such as Yeh et al 2020. The first is that it demonstrates that satellite data can accurately predict spatial variation in local headcount poverty rates, in addition to mean per capita consumption. This is important for informing policies that explicitly seek to target areas characterized by high rates of poverty. Second, this paper uses post-Lasso estimation to predict welfare and poverty rates instead of ridge estimation, which generates unbiased estimators of welfare and poverty. The third significant difference with the existing literature is that the measures of poverty and welfare used to train the prediction model are taken from model-based estimates of welfare and poverty rates derived from census data, rather than design-based estimates derived solely from household survey data. As a consequence, the headline results on predictive accuracy are not comparable to other papers in the literature that train models to cluster averages derived from household surveys (Jean et al. 2016; Yeh et al 2020). In the Sri Lankan context, model-based estimates of GN-division poverty and welfare derived from the census are more precise than design-based estimates from the sample, for two reasons. First, the model-based estimates leverage census auxiliary data that contain data on far more households in each GN Division than a typical household survey. Second, the model-based estimates of per capita consumption and poverty rates are based on averages over 100 draws for each household from the distribution of unexplained welfare, which is assumed to be stochastic. For poverty rates, this allows each household to be assigned an estimated probability of being poor, which prevents information loss that occurs when each household is dichotomously classified as either poor or non-poor no matter how far from the poverty line their measured welfare lies. For mean welfare, taking repeated simulations for each household virtually eliminates variation due to both unexplained shocks and measurement error in household welfare, and therefore provides a different indicator of longer-term predicted welfare. In the Sri Lankan context, the additional precision provided by incorporating both the full census data and 100 simulations per household substantially improves the predictive performance of the model. This illustrates that the explanatory power of prediction models, as measured by their R^2 , is highly sensitive to how precisely the “ground truth” is measured.

The fourth significant difference from most of the existing literature, with the notable exception of Ayush et al (2020), is the utilization of imagery features that are based either on recognizable objects or texture algorithms developed for computer vision applications. This method offers several advantages for the estimation of poverty rates. Interpretable features may provide a more transparent understanding of the

3 This is not to say the method outlined in this paper is necessarily better than a CNN approach. The two contexts are dissimilar and the authors cannot make general claims about performance. However, the method here appears to perform better for poorer households in the sample relative to the method in Jean et al. (2016) for the poorest households in their sample.

underlying factors that explain geographic variation in welfare in different contexts. Additionally, 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. A feature-based approach can easily be extended to alternative welfare indicators, such as headcount poverty rates measured at different thresholds, without the extensive retraining that is necessary for some end-to-end deep learning methods. Finally, separating the satellite-based feature engineering from the poverty modeling stage may be a more feasible processing pipeline for economists and statisticians tasked with generating small-area estimates of poverty and welfare.

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 presents the baseline models. Section 5 compares this method to a direct CNN approach or models using NTL features only. Section 6 examines the predictive power of these features with a household asset index, and examines robustness to spatially stratified cross-validation—that is predicting into novel areas. Section 7 concludes.

2. Data Description

The analysis is restricted to a sample area of approximately 3,500 km² in Sri Lanka. National coverage was not feasible due to the high cost and only partial availability of high-resolution imagery.⁴ The study sampled DS Divisions conditional on HSRI being available, drawing areas from urban, rural, and estate sectors.⁵ According to the 2012 census, population by sector in Sri Lanka is rural (77.4 percent), urban (18.2 percent), and estate (4.4 percent) (Sri Lanka Department of Census and Statistics 2012). Population by sector in the sample is rural (45.9 percent), urban (46.2 percent), and estate (7.8 percent).

2.1 Details on Satellite Imagery

Figure 1 depicts the coverage area of our satellite imagery over a map of Sri Lanka. The satellite imagery consists of 55 unique “scenes” purchased from Digital Globe (now Maxar), covering areas specified in the study’s sample area. 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 0.46 m², 0.41 m², and 0.61 m², respectively in the panchromatic band and 1.84 m², 1.65 m², 2.4 m² respectively in the multispectral bands. Preprocessing of imagery included pansharpening, orthorectification, and image mosaicking. Most imagery was captured in either 2011 or 2012, although some imagery from 2010 was also used.⁶

2.2 Details on Poverty Data

Ideally village poverty and consumption statistics would be generated directly from the 2012/2013 Household Income and Expenditure Survey (HIES), a detailed survey that measures the consumption patterns of 25,000 households on approximately 400 consumption items. The survey contains an average of 8.4 households per GN Division in the 47 sampled DS Divisions, making GN Division poverty estimates that would be derived directly from the HIES imprecise. The study therefore draws on the simulations that were used to generate official DS Division poverty estimates, which draw on the 2011 Census of Population and Housing (Department of Census and Statistics and World Bank 2015). The study used these simulations to generate poverty rate and welfare estimates at the GN Division level.

The methodology to derive the poverty estimates follows the traditional method employed by the World Bank (Elbers, Lanjouw, Lanjouw 2003). 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

4 These data are rapidly becoming more available and less expensive as companies such as Planet and DigitalGlobe expand their archives and launch newer, more precise satellites with more frequent revisit rates.

5 Sri Lanka classifies sectors as urban, rural, or estate. The estate sector refers to plantation areas of more than 20 acres with 10 or more residential laborers. Except for sample stratification, the estate sector is grouped with the rural sector.

6 More detail on the satellite imagery is provided in the supplementary online appendix.

Figure 1. Coverage Area of High Resolution Satellite Imagery



Source: Author's calculation using data derived from Digital Globe.
Note: Sample area shown highlighted in white.

both the census and the HIES. Sixteen random effect models were estimated in the HIES, corresponding to different groupings of provinces. The models estimated using the HIES data are:

$$\ln W_{ic} = \beta X_{ic} + \eta_c + \varepsilon_{ic} \tag{1}$$

Where W_{ic} is the welfare, or per capita expenditure, of household i in cluster c , X_{ic} is a vector of predictor variables common to the census and survey, η_c is a random cluster effect, and ε_{ic} is a household-specific error term, both of which are assumed to be normal. Feasible GLS is used to estimate the variance of the household-specific error term in order to account for heteroscedasticity. With estimates for β , the variance of η_c , and household-specific estimates of the variance of ε_{ic} in hand, welfare for each household is simulated 100 times. Households are considered poor in each simulation if their simulated welfare falls below the poverty line and the GN poverty rate equals the average of the poverty indicator across

simulations and households for each GN. The procedure is described in more detail in Department of Census and Statistics and World Bank (2015).

Similar methods can be used to derive the poverty gap, denoted as P_1 and defined as the product of the headcount poverty rate and the average relative shortfall of poor households (Foster, Greer, and Thorbecke 1984):

$$P_1 = \frac{1}{100 * N} \sum_{s=1}^{100} \sum_{i=1}^q \frac{Z - W_{ics}^*}{Z} \quad (2)$$

Where N is the number of households in the GN, q is the number of poor households in the GN, Z is the poverty line, and W_{ics}^* represents simulated welfare for household i in cluster c in simulation s . Therefore, in the baseline specification, the study averages poverty rates over the one hundred simulations. Later, as a robustness check, poverty rates are considered for only one imputation. The study calculates GN headcount poverty rates and poverty for two poverty lines: poverty line 1 at the 10th percentile of the national per capita consumption distribution, and poverty line 2 at the 40th percentile. This is equivalent to \$3.00 and \$5.13 per day respectively in 2011 PPP terms, which is higher than the global extreme poverty line in 2011 prices of \$1.90 per day.

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.⁷ In contrast, Elbers et al. (2008) find significantly less in Minas Gerais, Brazil. The effect of spatial heterogeneity on the results presented is unclear. The authors 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 that 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.

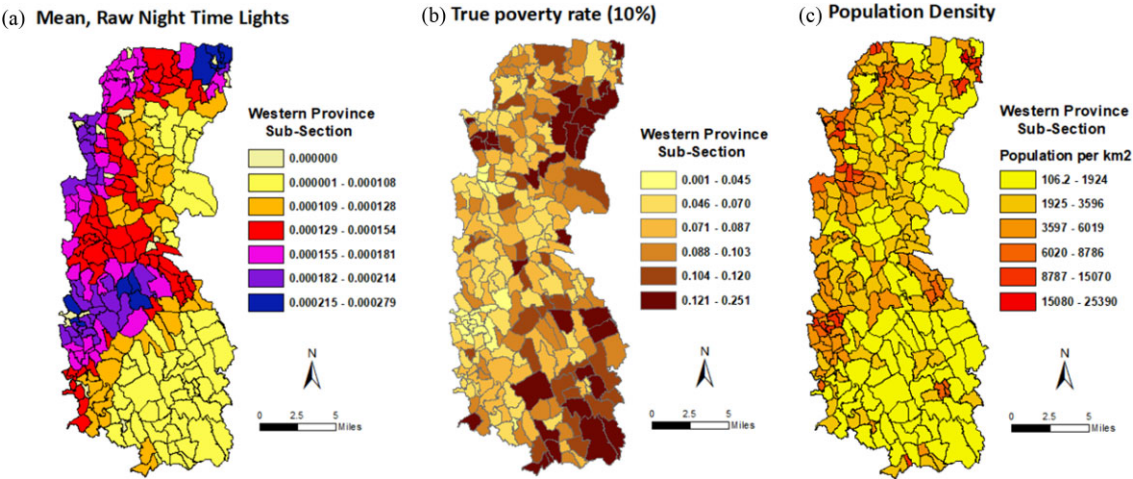
2.3 Comparison of GN Poverty Rates and Mean NTL Reflectance

A simple visual comparison between mean NTL and GN poverty rates illustrates why NTL provides limited information on subnational welfare. Figure 2 presents a panel of three images for the Western Province, Sri Lanka: mean raw NTL (left), poverty rates derived from the 10 percent national income threshold (middle), and log of mean population density (right). Comparing the left and middle panels, there is a modest association between villages that have low NTL reflectance and those that are high in poverty. Problems of overglow (Henderson, Storeygard, Weil 2012) could result in poor villages adjacent to wealthy ones being misclassified as nonpoor. 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 of only limited use for public policy.

The statistical correlation between GN NTL and population density is equally modest, at about 0.30. The study takes 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 an imperfect measure of welfare. The study therefore investigates whether the much richer set of information contained in HSRI daytime imagery translates into more accurate welfare predictions.

7 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.

Figure 2. Comparison of Mean Night Time Lights (NTL), Poverty Rate, and Mean Population Density, Western Province, Sri Lanka. (a) Average night time lights (NTL). (b) Average headcount relative poverty rate using 10th percentile of national income. (c) Population density



Source: Author's analysis based on 2012/13 Sri Lankan HIES, 2012 Census, and VIIRS NTL.
Note: Average headcount relative poverty rate using 10th percentile of national income.

2.4 Feature Extraction from High Resolution Satellites

The derived high-resolution spatial features fall into seven broad categories: (1) Agricultural Land, (2) Cars, (3) Building Density and Vegetation, (4) Shadows (building height proxy), (5) Road and Transportation, (6) Roof Type, and (7) Textural and Spectral characteristics. In addition to the satellite features, two geographic attributes of the GN Division are used: whether it is administratively classified as an urban area, and its area in square kilometers.⁸ Table 1 presents summary statistics for all variables.

Deep learning-based object classification was used for classifying 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 (proxy for building heights), and crop type. The classification method used is similar to Krizhevsky, Sutskever, and Hinton (2012), which utilizes convolutional neural networks (CNN) to build object predictions from raw imagery. Roof type, paved and unpaved roads of different widths, and railroads were classified using a combination of Trimble eCognition and Erdas Imagine software, utilizing a combination of support vector machines and visual identification. Classifier accuracy is greater than 90 percent for all of the objects recognized. Details on the extraction and classification process are provided in detail in the supplementary online appendix, which includes an example ROC curve for buildings.

2.4.1 Object Classification Details

The agricultural land variables consist 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 in subsequent regressions is GN Divisions with no agricultural land. The study also calculated the fraction of total GN area that is either paddy, plantation, or any agriculture. Figure 3 shows an example of a developed area building classification, with raw image shown at the top and CNN classification accuracy shown below. On the bottom panel, true positives are highlighted green, with false positives highlighted red. Figure 4 shows a sample car classification. Cars

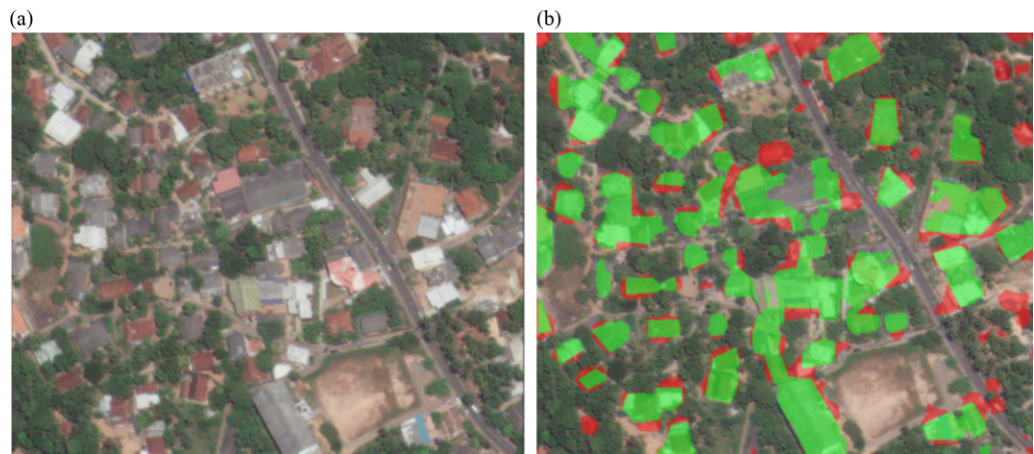
8 An urban indicator and area could in principle be calculated using remote sensing alone.

Table 1. Grama Niladhari Summary Statistics

	Mean	Sd	Min	Max
<i>Economic well-being</i>				
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
<i>Geographic descriptors</i>				
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
<i>Agricultural land</i>				
% 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
<i>Cars</i>				
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	0	0.00007	0	0.00093
<i>Building density and vegetation</i>				
% 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
<i>Shadows</i>				
ln shadow pixels (building height)	12.96	1.04	7.31	17.6
ln number of buildings	6.90	0.92	0	9.3
<i>Road variables</i>				
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
<i>Roof type</i>				
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
<i>Textural and spectral characteristics</i>				
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			

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

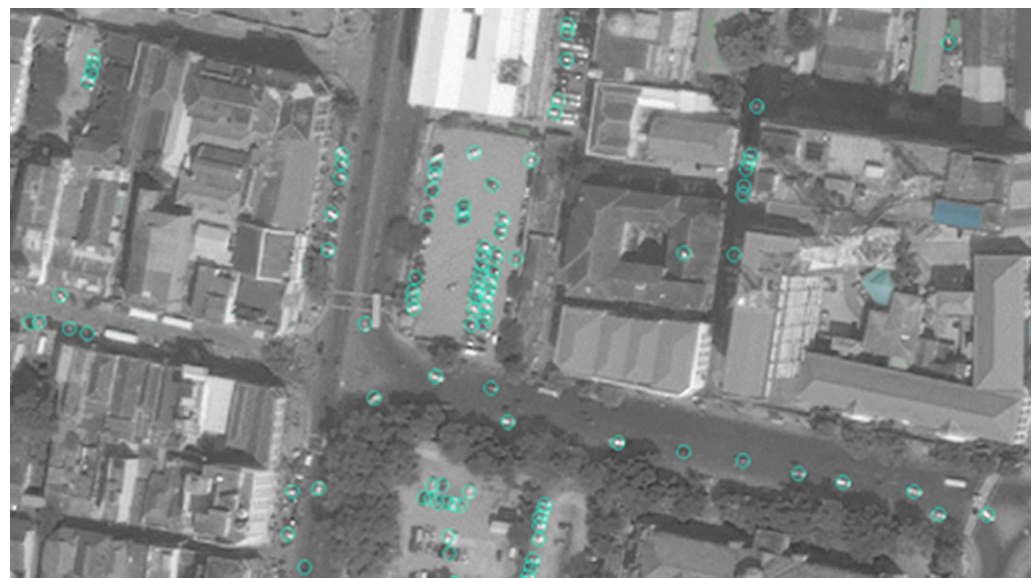
Figure 3. Example Developed Area (Buildings) Classification. (a) Raw satellite imagery. (b) Satellite imagery overlaid with building classification



Source: Digital Global.

Note: Areas in green show are true positive building classifications. Images in red are false positives: erroneously classified areas as buildings.

Figure 4. Example Car Classification



Source: Digital Globe.

Note: Cars identified by convolutional neural network shown in blue.

that are positively identified are shown circled in blue. False negatives are most prevalent where there is considerable tree masking of pixels.

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 the 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, the study takes the log of the car count, while imposing a smooth function for GNs with zero or few cars.⁹

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, and so forth—regardless of use or condition. These are grouped with two measures of the Normalized Difference Vegetation Index (NDVI). Although technically a spectral characteristic, the presence of vegetation in urban areas indicates development such as parks, trees, or lawns (i.e., areas that are 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 study includes two indicators that capture shadows of buildings: the log of the number of pixels classified as shadow as well as the fraction of shadows in a GN. The shadow variables use the angle of the sun as it shines on a building, and the shadows it displaces, to estimate the presence of tall buildings.¹⁰

The road variables that are calculated 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, the study calculates the fraction of roofs in a GN 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. Different roof materials exhibit different spectral properties, particularly in the subvisible bands of the spectrum. The roofs in the sample are clay (36.5 percent), aluminum (14.08 percent), asbestos (7.8 percent), or gray concrete (41.6 percent).

2.4.2 Details of Textural and Spectral Features (Contextual Features)

The study calculates seven separate types of contextual features: Fourier transform, Gabor filter, Histogram of Oriented Gradients (HoG), Line support regions (LSR), Pantex, and Speed-Up Robustness Features (SURF). These are often used in computer vision problems to decompose an image. 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. These features may be considered outputs from a dimension reduction technique, in that they are reduced dimensionality descriptions of a complex 2-D satellite imagery.

Because these measures may be novel to readers without backgrounds in remote sensing, further description may be helpful. The authors consider Pantex here to be a measure of human settlements. It is a spatial similarity index, where each cell is compared to adjacent cells in all directions. Open fields will have a low Pantex level, since cells in all directions have similar contrast, as will cells with straight roads. Dense cities 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. For more detail on imagery and the feature extraction process, see the supplementary online appendix.

3. Statistical Methodology

Given the list of available covariates, variable choice is not obvious. Estimating a model with the full set of candidate variables in table 1 would likely produce predictions that are overfit, in the sense that they

9 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.

10 Valid area refers to areas at the foot of building where shadows may appear.

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, bias is increased bias at the expense of lower variance.

The baseline model is a “Post-Lasso” estimator (Belloni and Chernuzhukov 2013). This two-step estimator first estimates a Lasso model over the full set of coefficients, followed by an OLS model over the set of non-zero coefficients from the Lasso step. The model that is estimated in the Lasso step is defined as

$$\beta_{Lasso} = \arg \min_{\beta} \left\{ \underbrace{\sum_{i=1}^N \left(y_i - \sum_{j=1}^K x_{ij} \beta_j \right)^2}_{\text{Residuals}} + \underbrace{\lambda \sum_{j=1}^K |\beta_j|}_{\text{Shrinkage factor}} \right\} \quad (3)$$

where the poverty rate in a GN is given by y_i and $\lambda \geq 0$ is a parameter that penalizes the absolute values of the coefficients. At the extreme, full relaxation of the penalization factor, that is setting λ to zero, yields unconstrained OLS estimates. Thus as $\lambda \rightarrow 0$, $\beta_{Lasso} \rightarrow \beta_{OLS}$. As $\lambda \rightarrow \infty$, the penalty increases and β_{Lasso} converges to the zero vector. Lasso regression is useful as a variable selection methodology because the sharp l_1 metric shrinks variables exactly to zero if they prove not to be useful in decreasing the sum of squared errors, thus creating a type of variable selection. However, simultaneously the Lasso “shrinks” the magnitude of coefficients towards zero, even for those that remain nonzero (Varian 2014). By subsequently estimating an OLS model for variables that remain nonzero after a Lasso model in the second stage, the study ensures that the coefficient estimates are unbiased (Belloni and Chernuzhukov 2013). To choose the appropriate value of λ , 10-fold cross-validation is applied, and the value of λ is chosen that minimizes root-mean squared error (RMSE) plus one standard error of estimated λ across folds.¹¹ GLM versions of the model, which ensures that predicted values lie in between zero and one, do not change the results qualitatively and are available by request.

Inferential standard errors are typically absent from Lasso models. Because of the Oracle property of the Lasso estimator (Fan and Li 2001), the standard errors from the OLS model are used in the second stage as the measures of population inference. 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 the study to use a single stage estimation strategy using only the selected variables present in the true data-generating process (Belloni and Chernuzhukov 2013).

4. Results

Table 2 presents the estimates from the main specification for the full sample. The first two columns show the model where GN poverty is defined at the lower poverty rate, the next two present the higher poverty rate models, and the next two present average GN consumption dependent variable models. Many extracted satellite features have high explanatory power, including agriculture type, length of roads and fraction of roads paved, number and density of buildings, NDVI, roof type, shadows (building height proxy), and two spatial features, LBPM, and Fourier transform. The models explain a high amount of the variation in poverty, summarized in the in-sample R-squared values between 0.608 and 0.618. Cross-validated R-squared, estimated using tenfold cross-validation, vary between 0.588 and 0.605. It is concluded from the results that the models are not likely to overfit to the data.

11 See Krstajic et al. (2014).

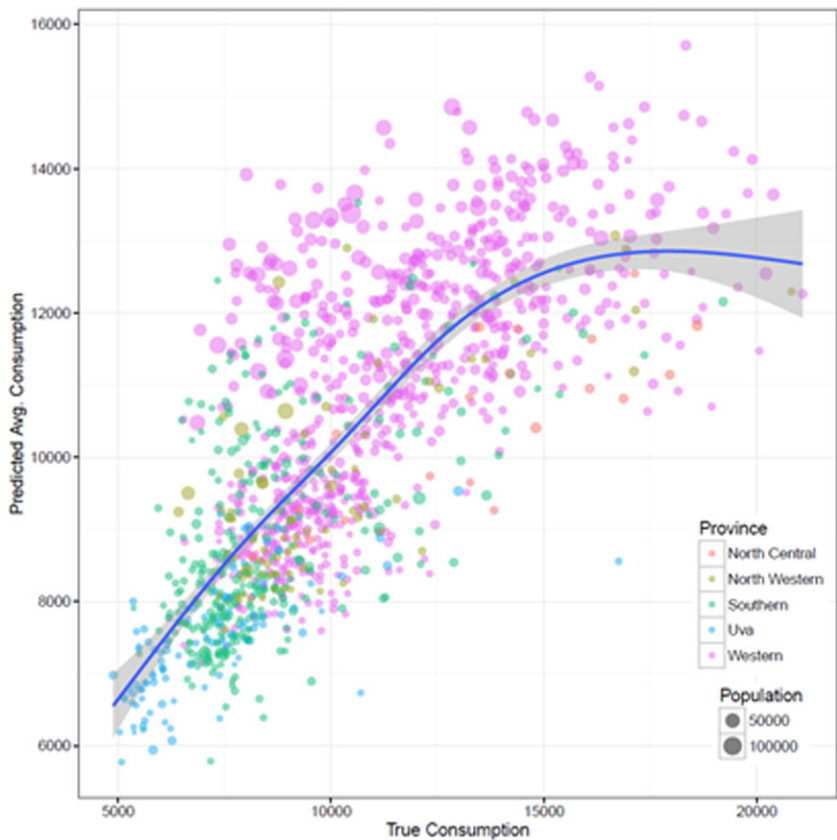
Table 2. Prediction of Local Area Poverty Rates Using High-Res Spatial Features

	Lower poverty rate (10% Nat. Inc.)		Higher poverty rate (40% Nat. Inc.)		Average 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) mean	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			
Cross-validated R-sq	0.588	0.605	0.594			
Cross-validated mean absolute error	0.032	0.078	0.139			

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.
Note: Unit of observation is Grama Niladhari (GN) division. Variables were selected using Lasso regularization from the candidate set of variables shown in [table 1](#).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Estimated standard errors may be biased downward due to the use of a common sample for model selection and estimation.

The results suggest that, in words, a simple linear model that includes only the geographic size of the GN Division, whether it is urban, and remotely sensed information explains 54–61 percent of the variation across GNs in headcount poverty rates. [Figure 5](#) plots predicted consumption against true average GN consumption, with colors assigned by province in which the GN is located. A Lowess smoothing line is shown with associated confidence interval. A perfect model would have predictions exactly on the 45° line. While there is noise, the predictions tend to straddle the 45° line indicating a high degree of agreement between the predicted and true welfare values, although the model tends to under-predict for wealthier GNs.

Figure 5. Model Diagnostic Plot of Predicted against True Average GN Consumption



Source: Author's analysis based on 2012/13 HIES, 2012 Census, and Digital Globe.
Note: Units are in 2012 Sri Lankan Rs.

4.1 Discussion of Significance and Magnitude of Satellite Features

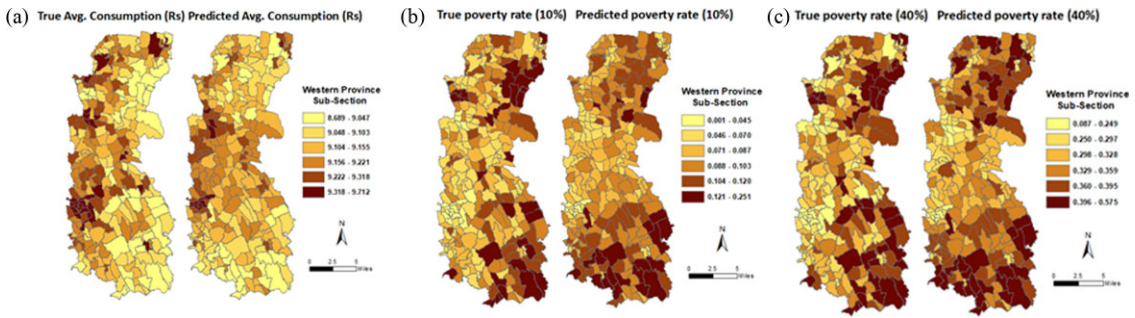
While the primary objective of this exercise is to obtain accurate predictions, the study also wants to shed light on which satellite features are most helpful in predicting poverty in this context. Additionally, table S2.1 in the supplementary online appendix presents estimated marginal coefficients.

The size of the GN, in square kilometers, is more strongly correlated with headcount or average consumption, but is statistically significant only for the lowest poverty rate model. This suggests that households in the bottom decile are disproportionately found in larger GN Divisions. The presence of agricultural land is weakly and negatively 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, Lasso selected three of the indicators for 10 and 40 percent poverty incidence models, and two for the log consumption model.¹² The results indicate a statistically significant negative relationship between the presence of paddy agricultural land and poverty, which is consistent with the relative deprivation of the tea plantation sector in Sri Lanka.

Compared with land type, the association between poverty and cars is mildly stronger, although not statistically significant in any of the specifications. Length of roads, fraction of roads paved, and runways

12 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.

Figure 6. Predicted versus True Welfare Measures, Average Consumption (top), 10 percent Poverty (middle) 40 percent Poverty (bottom). (a) Average household consumption. (b) Average headcount relative poverty rate using 10th percentile of national income. (c) Average headcount relative poverty rate using 40th percentile of national income.



Source: Author's calculations based on 2012/13 HIES, 2012 Census, and Digital Globe.
Note: Units are in 2012 Sri Lankan Rs.

are negatively associated with poverty, though only the first two are statistically significant, while GNs with more railways are poorer. Building density is strongly associated with log welfare and poverty and is statistically significant in all specifications. Vegetation is moderately associated with poverty and strongly statistically significant. For the lower poverty line model, both NDVI measures are selected. The higher poverty line 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, a proxy for building height, are selected—the share of valid area covered by shadows, and the log number of shadow pixels—and are statistically significant in most specifications. For roof type, the Lasso procedure selects both the fraction of roofs classified as clay and aluminum, for all three models and the post-Lasso model finds them strongly statistically significant. It also selects the fraction of roofs classified as asbestos for the lower poverty line model although this is not statistically significant. 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 concrete. 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 (Marx, Stoker, and Suri 2019).

Of the contextual features, 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 higher poverty line and log per capita consumption model. The coefficient on LBPM is strongly statistically significant in all of the specifications. The main exception is the mean of the Fourier transform, which is positively associated with poverty in the lower poverty line model, though the coefficient is not statistically significant. This is consistent with wealthier areas being laid out in a more orderly way, with more “right angles” in housing layouts.

Figure 6 presents a map showing the true welfare measures on the left panel, against the predicted welfare measures on the right, for Western Province, Sri Lanka. 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 lower poverty line. 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 underpredict poverty in the lowest poverty areas in the mid-west, suggesting that two-step or zero-inflated Poisson models may perform better.

Table 3. Shapley Decomposition of Share of Variance Explained (R^2) by High Resolution Spatial Feature Subgroup

	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)	Average log per capita consumption
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

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

Note: 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.

In summary, 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. Greater numbers of cars are 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 lush vegetation tend to be poorer. Clay and aluminum roofs, compared to grey roofs, are associated with greater levels of poverty. Of the contextual features, SURF exhibits a fairly strong association with poverty at the lower poverty line, 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.

4.2 Decomposition of Satellite Feature Explanatory Power

The results presented 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, the study decomposes the R^2 using a Shapley decomposition (Israeli 2007; Huettner and Sunder 2012; Shorrocks 2013). 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 candidate variables. 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 3 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 39 to 45 percent of the model's explanatory power. However, a number of other variables are moderately important. GN area, urban

Table 4. Urban and Rural Models of Local Area Poverty Rates (10 percent Relative Poverty Line) using High Resolution Spatial Features

	Rural		Urban	
	coef	t	coef	t
% of GN area that is agriculture	0.12*	[2.34]		
% of GN agriculture that is paddy	0.00076**	[3.11]	0.0002	[0.36]
log number of cars	0.019	[1.27]	0.085***	[5.73]
log area (square meters)	−0.033	[−1.43]		
log of sum of length of roads	0.029+	[1.93]		
fraction of roads paved	0.0012**	[3.44]	0.0014+	[2.06]
ln length airport roads	0.044***	[6.59]		
ln shadow pixels (building height)	−0.057**	[−3.23]		
Fraction of total roofs that are clay	−0.0041***	[−6.70]	0.0026	[1.41]
Fraction of total roofs that are aluminum	−0.0051***	[−5.63]	−0.0033+	[−1.84]
Fraction of total roofs are asbestos	−0.0017*	[−2.05]		
log of Total count of buildings in GN	0.040**	[3.53]	0.031	[0.77]
Vegetation index (NDVI), mean, scale 64	−0.27***	[−4.68]	0.28	[1.65]
Pantex (human settlements), mean	0.18***	[3.73]		
Linear binary pattern moments (scale 32m), mean	−0.013***	[−10.7]		
% of Total GN area that is plantation			−0.0058**	[−3.20]
ln length railroads			−0.0052	[−1.50]
% of area with buildings			0.028***	[6.07]
% shadows (building height) covering valid area			−0.015**	[−2.87]
Line support regions (scale 8m), mean			−1.4	[−0.34]
Fourier transform, mean			−0.0042+	[−1.96]
Constant	10.6***	[36.8]	8.89***	[27.6]
Observations	898		393	
R-sq	0.6562		0.4464	
R-sq adj.	0.6503		0.4274	
Cross-validated R-sq	0.5716		0.4184	
Cross-validated MAE	0.0327		0.02	

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

classification, road characteristics, roof type, and the texture 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.

4.3 Urban and Rural Linear Models

How does the relationship between indicators and welfare differ in urban and rural areas? Table 4 shows model estimates estimated separately for 393 urban villages and the 898 rural ones, based on Sri Lanka’s official definition of urban and rural areas.¹³ 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 15 for the rural model. R-squared values are slightly higher in rural areas (0.656) and significantly lower in urban areas (0.445).¹⁴ 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. The association between cars and poverty is significantly stronger in

13 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.

14 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, may better capture consumption in rural areas than in urban ones.

Table 5. Model Performance for Prediction of Average log per Capita Consumption at Different Points in the Welfare Distribution

	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
Cross-validated R-sq	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 p.c. income	8.83	8.95	9.00	9.09	9.16
Standard deviation	0.11	0.13	0.15	0.20	0.28

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

Note: 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 rightmost column is identical to the results reported in the right column of [table 2](#).

urban areas. In addition, 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 Model Performance at Varying Income Levels

The model's ability to predict variation in headcount poverty rates at both poverty lines suggests that it can effectively distinguish between households within lower parts of the welfare distribution. To verify this, the sample of GN Divisions is divided into quintiles based on the mean predicted per capita consumption of census households, and the main model for log per capita consumption on the subsample of the bottom 80, 60, 40, and 20 percent of the distribution is re-estimated. Model performance across income quintiles is shown in [table 5](#). Overall, the model continues to predict well within the poorest subsamples, as the R-squared declines from 0.60 in the full sample to 0.551 (in-sample) and 0.487 (cross-validated) when only considering the bottom decile. Given that the poorest decile of GNs have an average welfare of \$4.67 per day, this represents a little more than double the international poverty line. This suggests that this approach for estimating welfare from high-resolution satellites images is accurate for even moderately poor contexts.

4.5 Correcting for Spatial Autocorrelation

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, this study models explicitly the spatial autoregression (SAR) process and allows for SAR disturbances, a so-called SARAR model. This is implemented via a generalized spatial two-stage least-squares (GS2SLS) as shown in [Drukker, Prucha, and Raciborski \(2013\)](#). The results presented in [table 6](#) show that after correcting for spatial autocorrelation most high-resolution spatial features remain 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 Using Alternative Measures of Welfare as Ground Truth

The results so far have demonstrated that indicators derived from satellite imagery are strongly predictive of variation in welfare and poverty rates, using a measure of welfare that was simulated in the 2011

Table 6. MLE Estimation Correcting for Spatial Autocorrelation

	Average log per capita consumption	
	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	

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.
Note: Standard errors have been corrected according to Conley (1999), with model estimation via GMM. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

census. As the dependent variable, the baseline method uses, the average welfare or poverty rates, taken across both all census households in each GN Division and the 100 simulations of predicted residuals. This average is then regressed on various features derived from satellite data. Because the dependent variable is an average taken over 100 simulations, it is a measure of *expected* poverty and welfare across both simulations and GN households. When estimating poverty rates, this procedure uses the estimated variance of both the cluster and household idiosyncratic variance components to incorporate the full distribution of potential outcomes into the measure of poverty rates, which allows each household to be assigned an estimated probability of being poor rather than a dichotomous classification of poor or non-poor. Averaging over the one hundred simulations per household therefore reduces the variance of the estimated poverty rates and the measure of expected per capita consumption, which raises the explanatory power of the satellite indicators.

An alternative would be to compare satellite-based predictions against *simulated* poverty and welfare. One way to test this is to use the estimated GN mean log welfare and poverty rate for only one of the 100 simulations instead of the average across all simulations. This eliminates the additional precision obtained by averaging results for average welfare and poverty across 100 simulations of the stochastic distribution of unexplained welfare. In addition, because the simulation process is based on census data, it can be used to shed light on an additional question: How sensitive is the model’s predictive performance to the

Table 7. R^2 of Predicted Poverty and Welfare under Alternative Samples

	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)	Average log per capita consumption
Expected poverty rate and welfare over 100 simulations			
Full census sample	0.611	0.619	0.609
30 household census sample	0.548	0.581	0.551
8 household census sample	0.396	0.479	0.485
Poverty rate and welfare using one simulation			
Full census sample	0.373	0.469	0.486
30 household census sample	0.265	0.400	0.391
8 household census sample	0.167	0.273	0.348
Number of GN divisions		1291	
HIES subsample of GN divisions			
8 household census sample (one simulation)	0.174	0.319	0.378
HIES sample	0.217	0.259	0.322
Number of GN divisions		425	

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

Note: Cross-validated R^2 reported. Unit of observation is Grama Niladhari (GN) division. Independent variables are identical to those used in [table 2](#). Expected welfare refers to the average poverty rate or the average log per capita consumption averaged across both GN households and one hundred simulations.

number of households used to generate GN-level poverty and welfare estimates. In the results reported so far, the dependent variables—poverty and welfare—are calculated using simulations based on the full set of census data. On average, the census data contain approximately 500 households per GN Division. But census data are not typically available to calibrate models. In more typical settings, surveys are used that may contain 30 households per cluster, as in the Demographic and Health surveys. In Sri Lanka’s case, the Household and Income Expenditure Survey contains an average of eight households per GN Division in the sampled areas. When predicting log per capita consumption based on census characteristics, the average R^2 of the prediction models based on census characteristics is 0.46 ([Department of Census and Statistics 2015](#)). Therefore, increasing the average number of households in each GN by a factor of about 60 more than makes up for the fact that the results are based on a model, and generate far less noisy estimates. This in turn improves the predicted performance of indicators derived from satellite imagery.

[Table 7](#) shows the extent to which the predictive performance of the model depends on both the number of households and simulations used to generate the “ground truth” data used to train the model. The top row of [table 7](#) reports the in-sample R^2 when using expected welfare as the dependent variable, which is identical to the results reported in [table 2](#). Subsequent rows report R^2 s when fewer households are used to generate GN Division estimates, when one simulation instead of 100 is used per household, and when the sample is limited to GN Divisions present in the HIES survey. The bottom row of values shows R^2 s when using consumption as measured in the HIES sample. The set of satellite indicators used to predict poverty and welfare is the same for each row, matching those used for the main results shown in [table 2](#), so the only difference across rows is the dependent variable used in the regression.

Four main findings emerge from the table. First, the estimated R^2 value of the model is very sensitive to how the “ground truth” dependent variable is measured. Model R^2 values range from 0.17 to 0.61 for 10 percent poverty rates, from 0.26 to 0.62 for 40 percent poverty rates, and from 0.32 to 0.61 for log mean consumption. The highest R^2 s are obtained from the simulation model that averages across 100 simulations of the stochastic error terms for each household (expected welfare) and uses all household in the census. In contrast, the lowest R^2 s are obtained either by averaging over the observations in the HIES sample survey, or in the case of the 10 percent poverty measure, when using one simulation per household based on a census subsample containing eight randomly selected households per GN Division.

The second clear finding from [table 7](#) is that using one simulation rather than one hundred simulations per household significantly decreases the predictive power of the satellite indicators, particularly when predicting the 10 percent poverty rate. With the full census, the fall in R^2 is large, from 0.61 to 0.37. For the 40 percent poverty rate and mean consumption, the fall is noticeably smaller but still substantial, from 0.61 to 0.47 for 40 percent poverty and 0.61 to 0.49 for mean log per capita consumption. Using 100 simulations to estimate the probability that each household is poor is particularly beneficial when using lower poverty thresholds for generating accurate estimates of poverty rates, because it more accurately estimates the probability that households will be poor when their predicted per capita consumption, based solely on their census characteristics, lie just above the poverty line. For mean log per capita consumption, the higher R^2 when using 100 simulations per household reflects the greater correlation between satellite indicators and expected log per capita consumption, since virtually all of the noise added back by the simulation procedure is averaged away across the 100 simulations.

Third, the use of smaller subsamples of the census to estimate poverty and welfare explains the rest of the range of R^2 values. When using one simulation, the estimated R^2 for 10 percent poverty falls from 0.37 to 0.17 when using eight households per GN Division instead of the full census, and the comparable drop for 40 percent poverty rates is from 0.47 to 0.27. However, for mean consumption the drop in R^2 is smaller, from 0.49 to only 0.35. Finally, the explanatory power of the satellite variables is similar whether using the HIES sample itself or eight households per GN Division in the census sample and only one simulation. This remains true after limiting the estimation to the 425 GN Divisions included in the HIES sample. The only indicator in which the R^2 when using the sample exceeds the simulation with eight households per GN and one simulation is the 10 percent poverty rate, in which case the R^2 when using the sample is 0.22 and the R^2 for the simulated poverty rate is 0.17. This suggests that, to the extent that the measure of consumption collected in the HIES survey data contains accurate information on household transient shocks, these are not captured by the satellite data either.

This study's preferred estimates of GN-Division welfare and poverty rates remain the expected welfare and poverty rates based on averages across 100 simulations and the full set of census households, because these use all available information to generate local estimates of poverty and welfare. There are three important caveats that bear mentioning, however. The first is that expected welfare is a measure of predicted welfare that is largely free of measurement error and will not pick up transient shocks such as drought experienced by the village. It is therefore easier to predict using satellite imagery than average consumption taken from a typical household survey, as is reflected by the increase in R^2 from 0.49 to 0.61 when moving from one to one hundred simulations. A second caveat is that census-based imputations are not typically available to train prediction models, and if they are, there is no need to use geospatial data to generate predictions. However, these results do point to the importance of collecting high-quality training data. The type of training data that work well for design-based estimates may not be optimal for model-based estimates. For example, training data that cover a larger number of households in selected administrative units, or that collects welfare proxies such as assets for all households in an administrative unit to use in imputations, could be used to estimate models that fit the data better than a standard household survey.

The third and final caveat is that the estimated poverty rates from the model-based simulations contain a small amount of bias. This results from the assumption that the error terms in the model describing log per capita consumption follow a normal distribution, which is typical in small area estimation exercises ([Rao and Molina 2015](#)). However, as shown tables S5 and S6 in the supplementary online appendix, the bias introduced by the modeling procedure is small relative to the increase in precision. Therefore using the model-based estimates reduces the Mean Squared Error of the estimated mean welfare and poverty rates by about 90 percent. Using model-based estimates of estimated GN-Division poverty and log per capita consumption will, therefore, generate more accurate predictions of welfare and poverty rates using satellite data.

Table 8. Estimating Poverty Gap Using High Resolution 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	
Cross-validated R-sq	0.5855		0.6075	
Cross-validated MAE	0.0080		0.0282	

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

Note: Independent variables are the same as those listed in table 2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.7 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.¹⁵ This study computes the average poverty gap for each village, and uses 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. The study considers again poverty lines defined at the 10th and 40th percentiles of national consumption per capita. Table 8 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

15 The study calculates for its sample the FGT_1 metric (Foster, Greer, and Thorbecke 1984), which is defined as $FGT_1 = \frac{1}{N} \sum_{i=1}^N (\frac{z-y_i}{z})$, where y_i is an individual's income, and z is the poverty threshold.

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. Comparison to Alternative Estimation Methods

5.1 Comparisons to Convolutional Neural Networks

It is important to consider how the method used in this paper is different from other approaches to modeling poverty from satellite imagery, most notably the use of convolutional neural networks (CNNs) (LeCun et al. 1998). There is much similarity between a CNN used to model poverty directly, and the baseline method described in this paper. With a convolutional neural network, a series of filters is applied to images which produces a “feature map,” or outputs that highlight certain characteristics of the image. Using deep learning optimization methods, these filters are adjusted during training such that the model “learns” filters that are useful for the prediction task. The adjustment of several layers of filters is a very data-intensive task, often requiring millions of images to appropriately learn those that produce reliable feature maps for the specific prediction task. Researchers have used CNNs to directly model poverty both using transfer learning with Night Time Lights as an intermediate step (Jean et al. 2016) and using transfer learning using ImageNet weights (Babenko et al. 2017). Because poverty applications do not have access to millions of training examples, applications often use transfer learning, where weights are pre-defined from an auxiliary image prediction task. These auxiliary tasks often have millions of image examples, usually from the ImageNet data repository. Once these filters are constructed, they are then applied to the prediction task.

The baseline method used in this paper applies pre-built filters designed to recognize objects and other information from satellite images. Some of these—for the case of cars or building height—come from deep learning models. Others are filters used for specific remote sensing applications. The advantages of using pre-built satellite specific filters are: 1) The filters are specifically designed to recognize characteristics in satellite images rather than objects from still photography; 2) It may be more straightforward for these filters to incorporate additional information outside the visible spectrum¹⁶; 3) The satellite-specific features can be designed such that they carry interpretable information, such as number of cars or buildings. The disadvantage of using pre-built filters is that given that the filters are static, they cannot learn patterns unforeseen by the researchers that may be predictive of poverty.

To compare the prediction accuracy of this method with direct training via a CNN the study implemented a standard CNN model (ResNet-50) trained against the same imagery. ResNet-50 models have been used in a variety of computer vision tasks and are generally easy to train (Akiba, Suzuki, and Fukuda 2017). The study produced rectangular tiles of its satellite images at 250 by 250 pixels, producing 4130 images for training and 1044 for validation. The outcome variable—poverty of each GN—was discretized into bins of 0.05, for instance, [0 to 0.05), (0.05 to 0.1], and so forth. This was done because classification tasks are easier than regression tasks in CNNs. Data augmentation was further used to increase the number of imagery training samples. Two models were produced, one predicting poverty using a poverty line at the 10th percentile of national income, and one using a poverty line defined at the 40th percentile of national income. To optimize training, the study started with pre-trained ImageNet weights but re-optimized for 1000 epochs.¹⁷

16 Babenko et al. (2017) use an additional channel of infrared information from the satellite imagery. However, because weights from RGB channels are pre-trained starting at ImageNet weights, the model did not optimize to use the infrared channel. Yeh et al. (2020) use multispectral Landsat bands where the RGB bands are pre-trained to ImageNet weights, and the weights for the non-RGB bands in the first convolutional layer are set to the mean RGB channel weights. The difference in ease of training non-RGB bands may be due to the use of medium- and high-resolution imagery in the first study, and lower-resolution Landsat imagery in the second.

17 The model was implemented in PyTorch and is available at github.com/jonhersh/LKA_CNN_public.

Table 9. Comparison with CNN ResNet-50 Model

	Dependent variable	
	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)
Root mean squared error (RMSE), ResNet-50 CNN predicted poverty vs. true	0.0748	0.1762
Mean average error (MAE), ResNet-50 CNN predicted poverty vs. true	0.0554	0.1192
R^2 , ResNet-50 CNN predicted poverty vs. true	0.3949	0.2888

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.
Note: This table shows model diagnostics estimating a CNN ResNet-50 model against the same training data in the baseline model. Imagery was tiled into 250 by 250 pixel grids, which were processed through a ResNet-50 model in PyTorch. Model is pre-trained with ImageNet weights, with the final layers unfrozen and retrained for 1,000 epochs. Final model diagnostics reflect performance at the unit of analysis, Gram Niladhari Divisions (GNs).

Table 9 summarizes the results, showing in-sample prediction accuracy at the GN level. These results are comparable with results using the method outlined in this paper, if slightly worse. Results in the validation sample during training are roughly comparable to the in-sample results indicating a lack of overfitting. Cross-validation is not feasible with this study’s computational infrastructure given the large computational cost of training the model. For model of poverty with the poverty line defined at the 10th percentile of national income, the study estimates an MAE of 0.055 and an R^2 value of 0.3949. For the model with the poverty line at the 40th percentile of national income, the study estimates an MAE of 0.119 and an R^2 of 0.2888. The lower poverty line model produces results roughly comparable to the results in this study’s method, while the results with the higher poverty line model are slightly worse than this study’s baseline estimates.

The present article does not interpret this to mean that building models using pre-built filters necessarily dominates a direct CNN approach. However, given the current sample sizes for poverty training datasets, and existing CNN training methods, there does not appear to be a cost in terms of reduced prediction accuracy in using interpretable features. This is a conclusion that Ayush et al. (2020) also find. As dataset size increases, it’s very likely that this result will reverse, but researchers do not appear to be currently at that inflection point.

5.2 Comparisons to Night Time Lights

How does the predictive power of indicators derived from daytime imagery compare with night-time lights (NTL)? To shed light on this, table 10 presents OLS models covering the same sample area using NTL as the independent variable. The first three columns present poverty and per capita consumption models. Aggregate NTL 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 0.15 and 0.22, with performance lowest for the 10 percent head-count 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 three to four times as much variation in poverty or welfare as NTL. Columns (4)–(6) of table 4 show estimates that include DS Division fixed effects. Night-time lights are 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 NTL, one might also ask how much more explanatory power does NTL provide *in addition to* the indicators extracted from daytime imagery? Table 11 answers that question, by adding NTL to the Shapley decomposition. The NTL category includes average, squared, cubed, and average standard deviation of NTL. The NTL 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

Table 10. Model Estimates, Night Lights on Poverty/Average GN Consumption

	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)	Average log per capita consumption	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)	Average log per capita consumption
Avg night lights 2012	−98.47 [*] [−2.06]	−201.4 [−1.74]	262.3 [−1.23]	−26.82 [−1.40]	−58.45 [−0.99]	111.7 [−0.88]
Avg night lights squared	−16991.5 [−0.33]	−150061.7 [−0.98]	457585.9 [−1.47]	20678.3 [−0.73]	31038.3 [−0.34]	−54736.5 [−0.27]
Avg night lights cubed	26286044.1 [−1.5]	110121407.3 [−1.97]	−266793439.7 [*] [−2.31]	−3062554.7 [−0.38]	1662858 [−0.06]	−5719566 [−0.09]
Avg night lights std dev	0.0014 [−0.87]	0.00298 [−0.71]	−0.0049 [−0.62]	0.000408 [−0.51]	0.0000568 [−0.03]	0.00095 [−0.21]
Observations	1291	1291	1291	1291	1291	1291
R-sq	0.154	0.198	0.227	0.00485	0.00662	0.00758
R-sq adj.	0.151	0.196	0.225	0.00176	0.00353	0.0045
R-sq within				0.00485	0.00662	0.00758
R-sq between				0.324	0.407	0.472
R-sq overall				0.0868	0.118	0.133
Divisional Secretariat FEs	No	No	No	Yes	Yes	Yes

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.
Note: 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 11. Shapley Decomposition by High-Resolution Spatial Feature Subgroup and Night-Time Lights

	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)	Average log per capita consumption
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

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.
Note: Night time lights category includes the following transformations of night time lights: average, squared, cubed, and standard deviation. Variable groupings are identical to those in table 5.

that is captured through high-resolution satellite predictions. Furthermore, adding NTL 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 little explanatory power to the set of available daytime indicators.

6. Extensions and Applications

6.1 Correlation with Household Asset Index

Much of the existing literature that uses satellite data to predict household welfare uses household asset index values, typically estimated from a principal components analysis, as the dependent variable in the model. Asset indices are generally effective in ranking the welfare of households, especially in urban areas, but perform less well in rural households and identifying the extreme poor (Ngo and Christaensen 2019). Per capita consumption is also more responsive to shocks than asset indices are. Despite these shortcomings of asset indices, this study repeats the analysis using an asset index as the dependent variable, to shed light on how the nature of the dependent variable affects model fit.

The study constructs a measure of nonmonetary poverty derived from a principal component analysis, equal to the score of the factor loading of several individual welfare measures. This more transparently reflects observed welfare measures in the census. The welfare indicators and their associated factor loadings are listed in table S2.2 in the supplementary online appendix. They include six asset ownership dummy variables: home, computer, land phone, mobile phone, radio, and TV. The other variables included in the index measure are the quality of the housing floor and roof, the quality of sanitation facilities and services, the type of energy used for lighting and cooking, and the principal source of drinking water.²¹ The asset index is estimated based on census data from the 32 DS Divisions covered by the satellite imagery, and then averaged across all households in each GN Division. As in the per capita consumption and poverty modeling, post-lasso is used to select predictor variables from the full set of candidate variables.

The results of regressing the average asset index against the lasso-selected satellite predictors are shown in table 12. The third column reports the share of the R^2 explained by each different variable, according to a Shapley decomposition. Overall, the satellite features explain two-thirds of the variation in the average asset index at the GN Division level, a result quite similar to the CNN estimates reported in Yeh et al (2020). Of the variables in the regression, the variables that explain the most variation are those that are most closely related to population density. These include log geographic area, the urban dummy, built-up area, the seven spatial features, and building counts, each of which explains between 12 and 17 percent of the variation. The two NDVI measures, the two measures of shadows, and type of roof are less powerful, but each explains between 7 and 11 percent of the variation.

6.2 Poverty Estimation via Geographic Extrapolation

One motivation for using satellite imagery is to extrapolate poverty estimates into areas where survey data on economic well-being are not collected. 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 HIES failed to cover certain districts in the northern and eastern parts of the country due to civil conflict, and Pakistan's HIES exclude the Federally Administered Tribal Areas, Jammu and Kashmir.

To assess how well a model "travels" to a different geographic area, the study fits a series of models, where in each model it excludes a single Divisional Secretariat (DS), a larger administrative area, from the model, and uses the estimated model to predict into that excluded area. This is a form of "leave-one-out cross-validation" (LOOCV), a common method used to infer statistical out-of-sample performance (Gentle et al. 2012), where the unit of analysis is spatially stratified from distinct units. In that manner, this is similar to the spatially stratified cross-validation procedure recommended by Deville et al. (2014) where the study spatially stratifies according to distinct DS geographical units. In fact, it is identical when the number of partitions is equal to the number of Divisional Secretariats. In their context of population mapping using mobile phone data, Deville et al. (2014) find that failing to stratify at the geographic unit overstates model performance. The present study estimates both linear models and random forest

Table 12. Prediction of GN Division Average Asset Index Using High-Resolution Features

	Average asset index		Shapley contribution
	Coef	T	
log area (square meters)	−0.23**	−3.5	16.6%
= 1 if urban	0.45***	3.62	15.1%
Percent built-up	0.03*	2.41	17.1%
Percent shadow	−0.02	−1.91	5.9%
Log number of shadows	−0.04	−0.7	4.3%
Log of building count	0.28***	3.91	12.3%
Share clay roofs	0.00	−0.08	4.8%
Share aluminum roofs	−0.01**	−2.73	1.9%
Share asbestos roofs	0.00	−0.35	0.6%
Mean NDVI Scale 64	−0.49	−1.96	2.9%
Mean NDVI Scale 8	−0.17	0.14	4.4%
Pantex Scale 8	0.58***	0.14	
Hist of ordered gradients Scale 64	0.00***	0.00	
Linear binary pattern support Scale 32	−0.04***	0.01	
Line support region Scale 8	−17.43***	4.81	14.1%
Gabor Filter Scale 64	0.27	0.16	
Fourier transform Scale 32	0.00	0.00	
Speed-up robustness features Scale 32	0.03***	0.01	
R^2	0.686		
Cross-validated R^2	0.669		
Cross-validated MAE	0.364		
Number of observations	1291		

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.
Note: Unit of observation is Grama Niladhari (GN) Division. All models include a regression constant which is omitted from the table. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 14.1 percent in the right column refers to the combined Shapley contribution of the seven texture features listed between Pantex Scale 8 and Speed-up robustness features Scale 32.

models¹⁸ to predict out-of-sample to determine if more flexible model specifications perform better out-of-sample.

To give more detail on the spatially stratified leave one-one-cross-validation, the study starts with enumerating all Divisional Secretariats in the sample. It fits a model using all Divisional Secretariats except the first DS in the sample, and uses that fitted model to predict into the withheld DS. The study then withholds the second DS from estimation, fits a model using the remaining Divisional Secretariats, and uses that model to predict into the withheld DS. The study proceeds in this fashion until it has a prediction for each DS. Note that because Divisional Secretariats are large geographic units, this is a stronger test than non-spatially stratified sampling as the study may be predicting into areas that are geographically or economically distinct. This test is more representative of the intended application outlined.

Table 13 shows model performance at predicting into novel areas, comparing predicted and true welfare rates using both random forest and linear models to fit HRSF models. The novel area prediction error rates are larger than when predicting randomly out of sample using cross-validation. The study estimates R^2 values that vary between 0.488 and 0.579, which are smaller in magnitude than its baseline estimates, but not substantially smaller. While these error rates imply predicting into adjacent areas may be too imprecise for producing welfare measures intended as official statistics, they may be sufficient for generating rank ordering of villages by poverty or income.

18 For each random forest model the study uses 1000 decision trees, sampling 13 of the predictors with replacement.

Table 13. Divisional Secretariat Spatially Stratified Cross-Validation Model Performance

	Dependent variable		
	Average log per capita consumption	Lower poverty rate (10% Nat. Inc.)	Higher poverty rate (40% Nat. Inc.)
R^2 , Predicted and true poverty rates, linear models with full satellite variables	0.4876	0.4498	0.4586
R^2 , predicted and true poverty rates, random forest models with full satellite variables	0.5788	0.5643	0.5510
R^2 , predicted and true poverty rates, linear models with only night-time light variables	0.0652	0.0463	0.0510

Source: GN poverty statistics are derived from the 2012/13 HIES and 2012 Sri Lankan Census. Satellite features are derived from Digital Globe imagery.

7. Conclusion

Traditionally, given the prohibitive cost of conducting surveys sufficiently large to provide accurate statistics for small areas, generating small-area poverty estimates requires pairing a welfare survey with a census or inter-census survey. Census data are expensive to collect and are therefore produced relatively infrequently. Such data are also usually disseminated with a lag, making it difficult to rapidly assess changes in local living standards. The results show that indicators derived from high spatial resolution imagery, when paired with survey data, 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. 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? This study investigates these questions using a sample of 1,291 villages in Sri Lanka, linking measures of economic well-being with features derived from HRSI. The results indicate that the correlation between satellite-derived indicators and economic well-being is remarkably strong when using model-based measures of ground truth that use the full census data and average over one hundred simulations. Simple linear models explain 35 to 60 percent in the variation in poverty and average log per capita consumption. Models explain 68 percent of the variation in a household asset index. These models perform slightly better than an end-to-end CNN model trained over the same data, suggesting that models built with interpretable features do not come at a cost of predictive power.

Additional analysis also highlights the sensitivity of model performance to the measure of ground truth used for poverty and welfare. Predictive performance falls significantly, for example, when only using one simulation per household to estimate mean log per capita consumption and poverty rates at the GN Division level. In this case, simple linear models explain 37 percent of the variation in poverty rate when the poverty line is set at the 10th percentile, 47 percent of the variation in the poverty rate when the poverty line is set at the 40th percentile, and 49 percent of mean log consumption per capita. The explanatory power of the satellite indicators falls further when using only a subsample of census households to generate GN-Division poverty and welfare estimates, especially when predicting the 10 percent poverty rate. When using only 8 households per GN Division, the models explain only 17 percent of the variation in the 10 percent poverty rate, 32 percent of the variation in the 40 percent poverty rate, and 38 percent of the variation in mean log welfare.

The sensitivity of the predictive performance of models using satellite data to different measures of ground truth may have implications for the efficient design of sample surveys, when the surveys are

intended to be linked with geographically aggregated satellite data. Linking remote sensing and other “big data” with survey data can combine the decades of knowledge gained in collecting and interpreting survey data with the benefits of comprehensive big data. For example, as linking survey data with satellite imagery and other forms of big data become more popular, the benefits of collecting “micro-censuses” that interview all households in a random sample of low-level administrative areas may increase. At the same time, the sensitivity of models’ predictive performance to the precision of the training data highlights a challenge of using the R^2 as the sole measure of model performance. The models are intermediary inputs used to generate predictions, and the accuracy and precision of the predictions themselves ultimately matter more than the predictive performance of the models used to generate them. Estimating how targeting performance benefits from augmenting survey data with satellite indicators, and how this depends on the precision of the ground truth data used to train the model, is an important area for further research.

These findings raise a host of questions for further work. First, it is important to better understand the extent to which these results generalize to different social and 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 other features documented will hold in all environments. A second line of research could explore whether changes in satellite imagery could be used to forecast changes in economic well-being across space and time. Poverty surveys are typically collected every three years, and the most recent global estimates are produced with a three-year lag. 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. Third, there is much room for algorithm development, both for satellite feature extraction and model building. While this study has shown the direct CNN modeling to be no better than the study’s method, as the size of training datasets increases the CNN approach should overtake linear models.

Finally, and most pressingly, how should statistical agencies make best use of the wealth of information from satellite imagery? Given the increased quality and decreased cost of satellite images—and the continuing advancement of processing power to extract features and build models—how can statistical agencies adapt to make use of this information? Many of the features extracted—roads, number of buildings, amount of vegetation—would be useful as policy variables for additional uses beyond poverty mapping. Should statistical agencies develop these themselves, or should they rely on the many third-party agencies that supply this information? The conclusion of this study that if satellite features continue to be valuable, there is room for multilaterals to provide this information.¹⁹ The concern is that private companies will own these data pipelines and possibly extract excess surplus from their use (Hersh, Engstrom, Mann 2020).

Overall, the inevitable increase in the availability of imagery and feature identification algorithms, in conjunction with the encouraging results from this study and others in the literature, implies that satellite imagery will become an increasingly valuable tool to help governments and stakeholders better understand the spatial nature of poverty and economic welfare.

8. Data Availability

Due to the data sharing agreement signed between the authors and the Sri Lankan Department of Census and Statistics, the underlying poverty and income data cannot be shared publicly. Researchers may contact the Sri Lankan Department of Census and Statistics to obtain a data-sharing agreement: <http://www.statistics.gov.lk/ContactUs/headOffice>. Interested researchers may contact the corresponding author (hersh@chapman.edu) to facilitate the application for access to the data.

19 UN Global Pulse PulseSatellite is an excellent example of an open-data satellite analytics tool, but it focuses primarily on humanitarian efforts. <https://www.unglobalpulse.org/microsite/pulsesatellite/>.

References

- Akiba, T., S. Suzuki, and K. Fukuda. 2017. *Extremely Large Minibatch SGD: Training resNet-50 on ImageNet in 15 Minutes*. arXiv preprint arXiv:1711.04325. Cornell University, Ithaca, NY.
- Anselin, L. 2013. *Spatial Econometrics: Methods and Models*. Vol. 4. Springer Science & Business Media.
- Anselin, L., A.K. Bera, R. Florax, and M.J. Yoon 1996. "Simple Diagnostic Tests for Spatial Dependence." *Regional Science and Urban Economics* 26 (1): 77–104.
- Athey, S. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science* 355 (6324): 483–85.
- Athey, S., and G. Imbens. 2015. *Machine Learning Methods for Estimating Heterogeneous Causal Effects*. arXiv preprint arXiv:1504.01132. Cornell University, Ithaca, NY, USA.
- Ayush, K., B. Uzkent, M. Burke, D. Lobell, and S. Ermon. 2020. "Generating Interpretable Poverty Maps using Object Detection in Satellite Images." arXiv preprint arXiv:2002.01612.
- Babenko, B., J. Hersh, D. Newhouse, A. Ramakrishnan, and T. Swartz. 2017. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, with an Application in Mexico." Proceedings from NIPS 2017: Neural Information Processing Systems Workshop on Machine Learning for the Developing World. Long Beach, CA.
- Battese, G.E., R.M. Harter, and W.A. Fuller. 1988. "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data." *Journal of the American Statistical Association* 83 (401): 28–36.
- Belloni, A., and V. Chernozhukov. 2013. "Least Squares after Model Selection in High-Dimensional Sparse Models." *Bernoulli* 19 (2).
- Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of econometrics*, 92(1): 1–45.
- Dalal, N., and B. Triggs. (2005). "Histograms of Oriented Gradients for Human Detection." In *Computer Vision and Pattern Recognition (CVPR)*. 886–93. San Diego, CA.
- Department of Census and Statistics. 2012. "Sri Lanka Census of Population and Housing 2011."
- Department of Census and Statistics and World Bank. 2015. "The Spatial Distribution of Poverty in Sri Lanka." http://www.statistics.gov.lk/poverty/SpatialDistributionOfPoverty2012_13.pdf.
- Deville, P., C. Linard, S. Martin, M. Gilbert, F.R. Stevens, A.E. Gaughan, and A.J. Tatem. 2014. "Dynamic Population Mapping Using Mobile Phone Data." *Proceedings of the National Academy of Sciences* 111 (45): 15888–93.
- Donaldson, D., and A. Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives* 30 (4): 171–98.
- Drukker, D.M., I.R. Prucha, and R. Raciborski. 2013. "Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances." *The Stata Journal* 13 (2): 221–41.
- Elbers, C., J.O. Lanjouw, and P. Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality." *Econometrica* 71 (1): 355–64.
- Elbers, C., P.F. Lanjouw, and P.G. Leite. 2008. "Brazil Within Brazil: Testing the Poverty Map Methodology in Minas Gerais." World Bank Policy Research Working Paper Series, Vol.
- Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, and E.R. Davis. 1997. "Mapping City Lights with Nighttime Data from the DMSP Operational Linescan System." *Photogrammetric Engineering and Remote Sensing* 63 (6): 727–34.
- Engstrom, R., D. Pavelsku, T. Tomomi, and A. Wambile. 2019. "Mapping Poverty and Slums Using Multiple Methodologies in Accra, Ghana." *Joint Urban Remote Sensing Conference*, Vannes, France. May 22–24, 2019, 1–4.
- Engstrom, R., A. Sandborn, Q. Yu, J. Burgdorfer, D. Stow, J. Weeks, and J. Graesser. 2015. Mapping Slums Using Spatial Features in Accra, Ghana. Joint Urban and Remote Sensing Event Proceedings (JURSE). Lausanne, Switzerland, 10.1109/JURSE.2015.7120494.
- Engstrom, R., A. Copenhaver, D. Newhouse, J. Hersh, and V. Haldavanekar. 2017. "Evaluating the Relationship between Spatial and Spectral Features Derived from High Spatial Resolution Satellite Data and Urban Poverty in Colombo, Sri Lanka." Joint Urban Remote Sensing Event (JURSE 2017) Dubai, UAE. DOI: 10.1109/JURSE.2017.7924590
- Fan, J., and R. Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *Journal of the American Statistical Association* 96 (456): 1348–60.
- Foster, J., J. Greer, and E. Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–6.

- Gechter, M., and N. Tsivanidis. 2018. "The Welfare Consequences of Formalizing Developing Country Cities: Evidence from the Mumbai Mills Redevelopment." Working Paper. https://economics.yale.edu/sites/default/files/mumbaimills_ada-ns.pdf.
- Gentle, J. E., W.K., Härdle, and Y. Mori (Eds.). 2012. *Handbook of Computational Statistics: Concepts and Methods*. Berlin, Heidelberg: Springer-Verlag.
- Glaeser, E.L., S.D. Kominers, M. Luca, and N. Naik. 2015. *Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life* (No. w21778). National Bureau of Economic Research. Cambridge, MA, USA.
- Graesser, J., A. Cheriyyadath, R.R. Vatsavai, V. Chandola, J. Long, and E. Bright. 2012. "Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (4): 1164–76.
- Head, A., M. Manguin, N. Tran, and J. Blumenstock. 2017. "Can Human Development Be Measured with Satellite Imagery?" Article No. 8, ICTD '17: *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, November 2017.
- Henderson, J.V., A. Storeygard, and D.N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.
- Hersh, J., R. Engstrom, and M. Mann. 2020. "Open Data for Algorithms: Mapping Poverty in Belize Using Open Satellite Derived Features and Machine Learning." *Information Technology for Development* 27 (2): 1–30.
- Huettner, F., and M. Sunder. 2012. "Axiomatic Arguments for Decomposing Goodness of Fit According to Shapley and Owen Values." *Electronic Journal of Statistics* 6: 1239–50.
- Israeli, O. 2007. "A Shapley-Based Decomposition of the R-square of a Linear Regression." *Journal of Economic Inequality* 5 (2): 199–212.
- Jean, N., M. Burke, M. Xie, W.M. Davis, D.B. Lobell, and S. Ermon. 2016. "Combining Satellite Imagery And Machine Learning To Predict Poverty." *Science* 353 (6301): 790–4.
- Krstajic, D., L.J. Buturovic, D.E. Leahy, and S. Thomas. 2014. "Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models." *Journal of cheminformatics* 6 (1): 1–15.
- Krizhevsky, A., I. Sutskever, and G.E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems*. 1097–105.
- LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, 86(11): 2278–2324.
- Marx, B., T.M. Stoker, and T. Suri. 2019. "The Political Economy of Ethnicity and Property Rights in Slums: Evidence from Kenya." *American Economic Journal: Applied Economics* 11 (4).
- Mellander, C., J. Lobo, K. Stolarick, and Z. Matheson. 2015. "Night-time light data: A good proxy measure for economic activity?." *PloS one*, 10(10): e0139779.
- Ngo, D.K., and L. Christiaensen. 2019. "The performance of a consumption augmented asset index in ranking households and identifying the poor." *Review of Income and Wealth*, 65(4): 804–833.
- Pinkovskiy, M., and X. Sala-i-Martin. 2016. "Lights, Camera... Income! Illuminating the National Accounts-Household Surveys Debate." *Quarterly Journal of Economics* 131 (2): 579–631.
- Rao, J.N.K., and I. Molina. 2015. *Small-Area Estimation*. Hoboken, NJ: John Wiley and Sons, Inc.
- Sandborn, A., and R. Engstrom. 2016. "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., H. Uematsu, C. Wieser, N. Yoshida, and A. Dabalen. 2015. "Data Deprivation: Another Deprivation to End." Policy Research Working Paper 7252. World Bank, Washington, DC, USA.
- Shorrocks, A.F. 2013. "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value." *Journal of Economic Inequality* 11: 1–28.
- Tarozzi, A., and A. Deaton. 2009. "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas." *Review of Economics and Statistics* 91 (4): 773–92.
- Tibshirani, Robert 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.
- Varian, H.R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–27.
- Yeh, C., A. Perez, A. Driscoll, G. Azzari, Z. Tang, D. Lobell, S. Ermon, and M. Burke. 2020. "Using publicly available satellite imagery and deep learning to understand economic well-being in Africa." *Nature communications*, 11(1): 1–11.