David Newhouse (The World Bank Group), Ani Silwal (The World Bank Group), Ryan Engstrom (George Washington University), and Adane Bedada (George Washington University)

The Spatial Distribution of Poverty in Sri Lanka in 2016

Reliable poverty statistics derived from household surveys are typically only available for large administrative areas. Samples are typically too small to generate reliable estimates at lower administrative levels because surveys are costly. However, policymakers and aid agencies increasingly demand timely and granular poverty statistics to identify pockets of poverty and allocate scarce public resources within a country. Small area estimation techniques borrow strength from auxiliary data with greater coverage to generate more precise estimates of poverty for small geographic areas or demographic groups. Traditionally, this exercise relies on estimating a model in a household survey that predicts welfare based on household characteristics and applies the estimated coefficients and parameters to census data to simulate poverty rates.

In Sri Lanka, previous poverty map exercises conducted for 2002 and 2012/13 were powerful tools in measuring and comparing poverty at sub-district levels. These maps were used by policy makers during the reform of the Samurdhi transfer program and during the implementation of other programs aimed at reducing poverty. Despite their usefulness in guiding policies to reduce poverty, the 2012/13 poverty map has become outdated and may no longer reflect recent developments in Sri Lanka's economic conditions such as the end of the civil conflict in 2009 as well as public investment in infrastructure and social protection. A new poverty map could inform policy makers about whether previous pockets of poverty have persisted and whether new pockets have emerged. The goal of this exercise is to generate small area estimates of poverty in Sri Lanka in 2016 that incorporates the latest available data and methodological advances. The major challenge is that the census and the survey were conducted four years apart, making it unrealistic to assume that household characteristics in the survey and the census are drawn from the same statistical distribution. Therefore, it is not appropriate to directly apply the methodology proposed in Elbers et al. (2003, also known as the ELL model) and used in the 2002 and 2012/13 poverty maps.

We address the temporal gap between the census and the survey by excluding household characteristics from the consumption module. We instead use geographic-means of variables from the census, following Lange, Pape, and Wollberg (2018), as well as remote-sensing data. Because these indicators are linked with the household survey geographically, they can be used both to develop the prediction model and simulate welfare in the census. In other words, we build a model that predicts household consumption, taken from the HIES 2016 data, as a function of geographic means of census and remote census variables. The model coefficients and parameters are then used to predict consumption for all census households. The variables used in the model stage are therefore the same variables, drawn from the same distribution, as those used in the simulation stage. One consequence of this method is that the explanatory power of the consumption model is lower than a model that incorporates household characteristics, because the explanatory variables can only capture variation across villages. As a result, sub-district estimates obtained when excluding household predictors are less precise, and it is crucial to carefully evaluate the precision and reliability of the estimates using this method.

Satellite imagery is an exciting new source of data for constructing poverty maps. The availability of satellite imagery and the ability to process it has improved significantly in recent years. Satellite imagery has extensive spatial and temporal coverage, especially for recent years, and captures a significant portion of spatial differences in living standards across villages (Jean et al, 2016, Engstrom et al, 2017). We also use data on rainfall shocks, night lights, and built-up area, in addition to features derived from satellite imagery. This paper applies a variety of variables derived from remote sensing data and the 2011 census, linked with the 2016 household survey at the sub-district level, to generate sub-district poverty estimates for Sri Lanka in 2016. The empirical model links household per capita consumption to village characteristics using the 2016 Household Income and Expenditure Survey. Village characteristics are selected from a pool of census and satellite indicators using the Least Absolute Shrinkage and Selection Operator (LASSO) to avoid overfitting. The selected variables are then employed in a simulation framework that incorporates random location effects, a GLS heteroscedasticity correction, and Empirical Bayes estimation (Van der Weide, 2014, Nguyen et al, 2018).

Incorporating geographically linked satellite indicators reduces the average coefficient of variation of subdistrict poverty estimates by approximately one third compared with direct survey-based estimates. This is roughly equivalent to doubling the effective size of the sample. The coefficient of variation (CV) of augmented sub-district estimates of mean per capita consumption is 10.3%, about a third lower than the CV of the direct estimates that can be obtained from survey data (17.9%). This is also close to the survey's 10% percent target for district level estimates, which are routinely published. The analysis compares the performance of the estimates derived from satellite indicators with the traditional method of applying a model estimated on 2016 survey data to 2012 census data. Not surprisingly, predictions derived from satellite and census data generate much more plausible estimates of changes in sub-district poverty rates. Models estimated using either census data alone, or satellite data alone, offer similar predictive performance to models that draw on both types of variables. Results from a cross-validation exercise suggest that the Mean Squared Error of our results is much lower than that of the traditional ELL method (0.18 vs. 0.26) as is the Mean Absolute Deviation (2.58 vs. 3.35). Overall, the results strongly suggest that indicators derived from satellite imagery or other geographically linked data should routinely be combined with survey data to better track changes in poverty at local areas, even in the absence of a new census.