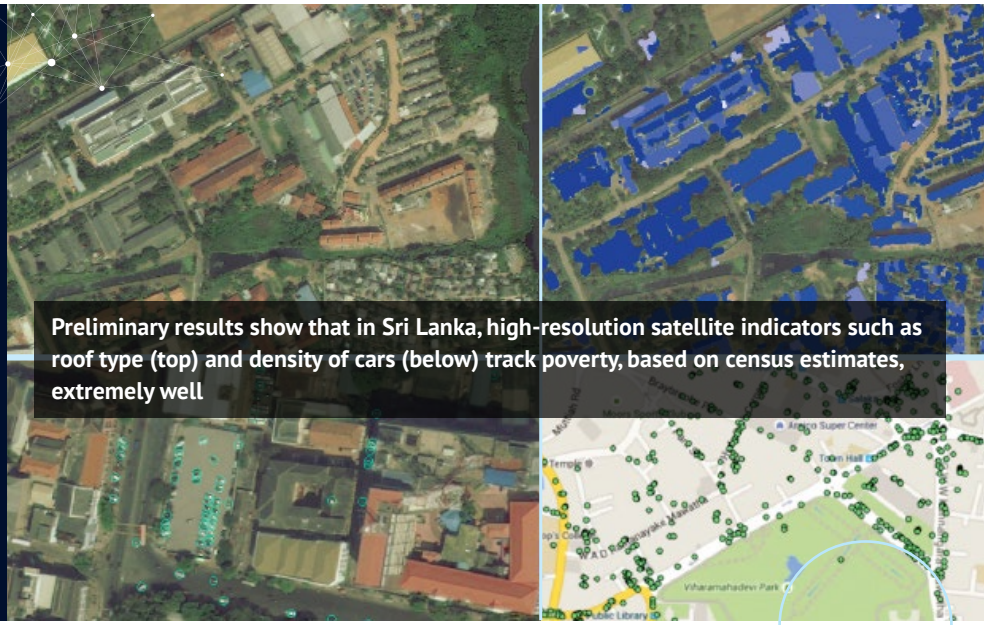
BIG DATA
CASE Study

WORLD BANK GROUP



Preliminary results show that in Sri Lanka, high-resolution satellite indicators such as roof type (top) and density of cars (below) track poverty, based on census estimates, extremely well

Mapping Poverty by Satellite

David Newhouse

SUMMARY

Poverty must be located accurately if development interventions are to be effectively targeted and monitored. However, expensive data collection and processing mean national poverty estimates in developing countries are often outdated. In response, this project explored the use of indicators derived from satellite data to predict geographic variations in poverty. The first component examined how well publicly available low-resolution satellite indicators such as nighttime lights and land type contribute to poverty prediction, and how this depends on the method used to build the prediction model. When satellite indicators were applied to the models in Pakistan, which is unusual in that it can generate district poverty estimates from a detailed household survey, they did not improve the accuracy of predictions. The rich survey information meant satellite-based indicators contributed nothing new. However, in Sri Lanka, which is more typical in generating poverty estimates from a census (meaning fewer indicators), even freely available satellite indicators improved the accuracy of predictions. In both cases, the team found models selected using an innovative statistical technique called Lasso to work best for predicting poverty at a more local level, with sizeable benefits when there are many variables.

When high-resolution satellite data indicators – such as cars, built-up area, shadows, roof type and road type – were combined with Sri Lankan census data, preliminary results showed that satellite indicators track regional differences in poverty extremely well. They demonstrate that high-resolution satellite imagery is a valuable complement to household survey data, with potential to help generate more accurate and updated local poverty maps and refine targeting in development initiatives.

CHALLENGE

Development interventions can be more effectively targeted and monitored if poverty can be located more precisely. However, long lags in processing and report-writing mean that recent national estimates of poverty in developing countries are often several years old. In addition, local-level estimates require census data that is expensive and collected infrequently. Big data, in the form of satellite imagery, has so far been largely untapped by policymakers wanting to understand where exactly the poorest people live. Little is known about which satellite-based indicators help predict poverty, and there is uncertainty around the best way to build a prediction model. Although numerous models have been developed, there has been little rigorous comparison of different approaches.

This project aimed to assess different approaches to poverty mapping, as well as the extent to which high-resolution satellite imagery can be used to generate more accurate poverty estimates. Incorporating satellite data into poverty mapping is a first step towards using the wealth of non-traditional data generated daily to predict poverty more effectively. Satellite-based data analysis is particularly attractive because it can see a complete picture of a particular area, unlike, for example, mobile phone-based analysis, which typically captures only a subset of phone users. Satellite data can also be collected frequently at fine geographic levels, even in conflict areas not conducive to surveys. Using this data to better understand poverty can help development practitioners target interventions and evaluate their effectiveness more accurately. Satellite-enhanced maps would be a key step towards the goal of real-time estimates of how pockets of poverty are evolving.

Satellite imagery gives insights into factors such as the scale of urbanization, infrastructure and natural resources

INNOVATION

The study involved two stages:

- The first compared different methods of generating poverty prediction models, in both Pakistan (which has existing data for many poverty indicators) and Sri Lanka (where fewer are available).
- The team then examined how high-resolution satellite data indicators are correlated with poverty predictions based on the 2011 Census, for local administrative areas in Sri Lanka.

Assessing poverty prediction models

To examine different approaches to developing poverty prediction models, the team applied out-of-sample validation techniques to household data from Pakistan and Sri Lanka. A randomly selected portion of the sample was repeatedly withheld when generating the prediction model, and accuracy was assessed by comparing extrapolated poverty rates from the prediction

model to actual poverty rates in those withheld areas. This technique was used to compare the accuracy of models derived using manual selection, stepwise regression and Lasso-based procedures. The team also augmented the set of prediction variables with publicly available low-resolution satellite data to see whether this improved traditional poverty mapping techniques.

Applying higher-resolution imagery

In the second stage, the team purchased high-resolution (0.5m per pixel) satellite imagery covering approximately 5 percent of Sri Lanka, including both rural and urban areas, and containing roughly 1,400 local administrative divisions. They used multispectral imagery (multiple images taken at varying spectrum wavelengths) to capture variations in roof texture and surface material, enabling far more accurate identification of possible correlates of income. Novel methods are also emerging to detect smaller objects such as cars from such imagery. These additional predictors have not yet been tried in poverty mapping models. Because some indicators, such as car traffic, can change rapidly as economic growth occurs, these additional predictors could pave the way towards more frequent poverty estimates in the future. Contractors produced pan- sharpened mosaics (merging several smaller scenes) of the raw high-resolution imagery. The team then worked with experts to develop detection algorithms to identify possible poverty predictors that can be extracted from high- resolution satellite data. These include built-up area, building and car density, type of roofing, amounts of shadow, road type and agricultural land-use. Using open-source image processing algorithms, the team also calculated whether buildings were more rectangular or had more chaotic angles (indicating higher poverty) and constructed indicators such as the share of paved roads or built-up area. They assessed the density of each feature in each local district division and correlated the satellite-based measures with poverty estimates from the 2011 census data. This provided a measure of which indicators correlate most strongly with predicted poverty and other measures of economic welfare from the census. The process demonstrates the potential of high-resolution satellite data to capture new indicators correlated to poverty.

RESULTS

The assessment of poverty prediction models showed that the number of poverty indicators affects the performance of different models and the usefulness of adding satellite data to the indicators. In Pakistan, with many potential indicators, the team found that Lasso models outperform both discretionary and stepwise models. However, Lasso and stepwise models give comparable results in Sri Lanka, where the set of indicators is smaller. The accuracy of the prediction model also depends considerably on the poverty threshold. In Sri Lanka, models were better able to predict the bottom 40 percent than the bottom 10 percent, but in Pakistan the reverse was true. In Sri Lanka, including publicly available satellite data made poverty predictions more accurate, but in Pakistan, the satellite data makes predictions slightly less accurate. When the satellite data is included in Sri Lanka, the Lasso models significantly outperform the manual and stepwise models.

Overall, the team found Lasso-based models are preferred for generating poverty predictions, and that the benefits can be sizeable when the pool of candidate variables is large, as in the case of Pakistan and of Sri Lanka when satellite indicators are included. There is strong interest in testing different poverty modeling approaches and the research highlighted the value of using publicly-available satellite data to generate small-area estimates of poverty in contexts where census data is limited.

Understanding pictures of poverty

For the Sri Lanka study using high-resolution imagery, preliminary results show that indicators track regional differences in poverty, based on estimates from the census, extremely well. When looking at the full sample, the key indicators relate to building density and urbanization, including the number of buildings, a vegetation index, and in rural areas, the share of roads that are paved, shadow, and type of roof. But when predicting variation in poverty in local areas within urban areas, the number of cars and an abstract measure of rectangular buildings become strong predictors as well.

These preliminary results demonstrate that satellite-based data is a valuable complement to household survey data, strengthening the case for investing in high-resolution imagery to monitor poverty more generally, as well as project impacts. This approach is the first step of an exciting research agenda. Imagery can deliver new insights related to a variety of development challenges, such as the scale of urbanization, infrastructure and the state of natural resources. Much more work is needed to explore which indicators best track local variations in poverty in a variety of contexts. These might include building density, roads, agricultural land or forest cover. More analysis is also needed to better understand the tradeoff between the quality and cost of the imagery on the one hand, and its benefits in terms of predicting local variation in poverty. Eventually, satellite-based imagery could also be a valuable tool in improving measuring inequality, monitoring development projects and 'nowcasting' poverty rates.

LESSONS LEARNED

As the price of high-resolution imagery continues to fall and coverage improves, satellite-based data will become an increasingly useful source of information about welfare in developing countries

- ***Drive research into mainstreaming the use of satellite imagery in poverty measurement***
  Most poverty economists are unaware of satellite technology's potential to improve small area estimates. There is not yet sufficient evidence to mainstream the use of satellite imagery to improve poverty maps based on census data, but this project has increased awareness. Beyond poverty prediction, satellite imagery can help efforts to better understand poverty. For example, road network estimates can indicate whether new road construction benefits the poor.
- ***Allocate sufficient resources to sourcing raw satellite imagery***
  Navigating the market for high-resolution satellite imagery and developing relationships with vendors and processing experts can take longer than anticipated. Ensure adequate project funding and planning to allow for the time this can take.
- ***Watch for future potential in satellite imagery***
  The accuracy and timeliness of satellite-derived poverty maps will continue to improve. More work is needed to see which satellite-based measures best predict poverty, both spatially and across time. There is broad scope for fruitful collaboration between poverty economists and geo-spatial image experts.